

THE IMPORTANCE SAMPLING TECHNIQUE FOR UNDERSTANDING RARE EVENTS IN ERDŐS-RÉNYI RANDOM GRAPHS

SHANKAR BHAMIDI¹, JAN HANNIG², CHIA YING LEE³, AND JAMES NOLEN⁴

ABSTRACT. In dense Erdős-Rényi random graphs, we are interested in the events where large numbers of a given subgraphs occur. The mean behaviour of subgraph counts is known, and only recently were the related large deviations results discovered. Consequently, it is natural to ask, what is the probability of an Erdős-Rényi graph containing an excessively large number of a given subgraph? Using the large deviation principle, we study an importance sampling scheme as a method to numerically compute the small probabilities of large triangle counts occurring within Erdős-Rényi graphs. The exponential tilt used in the importance sampling scheme comes from a generalized class of exponential random graphs. Asymptotic optimality, a measure of the efficiency of the importance sampling scheme, is achieved by the special choice of exponential random graph that is indistinguishable from the Erdős-Rényi graph conditioned to have many triangles. We show how this choice can be made for the conditioned Erdős-Rényi graphs both in the replica symmetric phase and also in parts of the replica breaking phase. Equally interestingly, we also show that the exponential tilt suggested directly by the large deviation principle does not always yield an optimal scheme.

1. INTRODUCTION

In this paper we study the use of importance sampling schemes to numerically estimate the probability that an Erdős-Rényi random graph contains an unusually large number of triangles. Consider an Erdős-Rényi random graph $\mathcal{G}_{n,p}$ on n vertices with edge probability $p \in (0, 1)$. For a simple graph X on n vertices, let $T(X)$ denote the number of triangles in X . For p fixed, one can show that $\mathbb{E}[T(\mathcal{G}_{n,p})] \sim \binom{n}{3} p^3$ as $n \rightarrow \infty$. For $t > p$, what is the probability

$$\mu_n = \mathbb{P} \left(T(\mathcal{G}_{n,p}) \geq \binom{n}{3} t^3 \right) \quad (1.1)$$

that $\mathcal{G}_{n,p}$ has an atypically large number of triangles? The last few years have witnessed a number of deep results in understanding such questions on upper tails of triangle counts, along with more general subgraph densities (see e.g., [3, 6–9, 13, 16]). In the dense graph case, where the edge probability p stays fixed as $n \rightarrow \infty$, [7] derived a large deviation principle (LDP) for the rare event $\{T(\mathcal{G}_{n,p}) \geq \binom{n}{3} t^3\}$, showing that for t within a certain subset of $(p, 1]$,

$$\mathbb{P} \left(T(\mathcal{G}_{n,p}) \geq \binom{n}{3} t^3 \right) = \exp \left(-n^2 I_p(t) (1 + O(n^{-1/2})) \right) \quad (1.2)$$

2010 *Mathematics Subject Classification.* Primary: 65C05, 05C80, 60F10.

Key words and phrases. Erdős-Rényi random graphs, exponential random graphs, rare event simulation, large deviations, graph limits.

where the rate function $I_p(t)$ is given by

$$I_p(t) = \frac{1}{2} \left(t \log \frac{t}{p} + (1-t) \log \frac{1-t}{1-p} \right). \quad (1.3)$$

More recently [8] showed a general large deviations principle for dense Erdős-Rényi graphs, using the theory of limits of dense random graph sequences developed recently by Lovasz et al. [3, 14, 15]. When specialized to upper tails of triangle counts, the large deviation principle shows that for the range of (p, t) considered in (1.2), the Erdős-Rényi graph $\mathcal{G}_{n,p}$ conditioned on the rare event $\{T(\mathcal{G}_{n,p}) \geq \binom{n}{3} t^3\}$ is *asymptotically indistinguishable* from another Erdős-Rényi graph $\mathcal{G}_{n,t}$ with edge probability t , in a sense that the typical graphs in the conditioned Erdős-Rényi graph resembles a typical graph drawn from $\mathcal{G}_{n,t}$ when n is large. (Asymptotic indistinguishability is explained more precisely at (2.11).) While this seems plausible for any $t > p$ since $\mathbb{E}[T(\mathcal{G}_{n,t})] \sim \binom{n}{3} t^3$ as $n \rightarrow \infty$, it is not always the case. Depending on p and t , it may be that the graph $\mathcal{G}_{n,p}$ conditioned on the event $\{T(\mathcal{G}_{n,p}) \geq \binom{n}{3} t^3\}$ tends to form cliques and hence does not resemble an Erdős-Rényi graph. When the conditioned graph does resemble an Erdős-Rényi graph, we say that (p, t) is in the *replica symmetric* phase. On the other hand, when the conditioned graph is not asymptotically indistinguishable from an Erdős-Rényi graph we say that (p, t) is in the *replica breaking* phase. (See Definition 2.2.)

Our approach to this problem is from a computational perspective: we study the use of importance sampling schemes for numerically estimating the probability μ_n , and also determine the schemes that perform optimally for those (p, t) in the replica symmetric phase as well as in a subset of the replica breaking phase.

The exponential decay of the probability of the event of interest makes it difficult to estimate this probability even for moderately large n . Direct Monte Carlo sampling is obviously intractable. The central strategy of importance sampling is to sample from a different probability measure, the *tilted measure*, under which the event of interest is no longer rare; one obtains more successful samples falling in the event of interest but each sample must then be weighted appropriately according to the Radon-Nikodym derivative of the original measure against the tilted measure. Importance sampling techniques have been used in many other stochastic systems, such as SDEs and Markov processes and queuing systems, see e.g [2, 4, 10, 12, 20] and the references therein. In particular, when a large deviations principle is known for the stochastic system, the tilted measure commonly used is a change of measure arising from the LDP. However, not every tilted measure associated with the LDP works well. It is well known that a poorly chosen tilted measure can lead to an estimator that performs worse than Monte Carlo sampling, or whose variance blows up [11]. Thus, a careful choice of tilted measure is of utmost importance.

Given (p, t) , the $\mathcal{G}_{n,t}$ -measure works as a tilted measure by ramping up the edge probability of the samples; we shall refer to $\mathcal{G}_{n,t}$ as an *edge tilt*. As we will see later on, even when the LDP suggests that $\mathcal{G}_{n,t}$ is asymptotically equivalent to the conditioned $\mathcal{G}_{n,p}$ graph, the edge tilt is not necessarily a good tilted measure for estimating the probability μ_n . It turns out that the class of measures associated with the Erdős-Rényi graphs is too limited, so we must broaden the class to consider the class of exponential random graphs. Exponential random graphs are generally defined via a Gibbs measure. In the context of estimating rare events for triangles, one need only consider the Gibbs measures involving only edge and triangle counts. Hence, consider the exponential random graphs $\mathcal{G}_n^{h,\beta,\alpha}$ defined via

the Gibbs measure, $\mathbb{Q}_n = \mathbb{Q}_n^{h,\beta,\alpha}$ on the space of simple graphs on n vertices, where

$$\mathbb{Q}_n(X) \propto e^{H(X)}, \quad \text{where } H(X) = hE(X) + \frac{\beta}{n} \left(\frac{n^3}{6} \right)^{1-\alpha} T(X)^\alpha$$

with parameters $h \in \mathbb{R}$ and $\beta, \alpha > 0$. $E(X)$ is the number of edges in graph X . Given (p, t) , a special choice of Gibbs measure $\mathbb{Q}_n^{h,\beta,\alpha}$ is what we will call a *triangle tilt*, which works by ramping up the probability of triangles. We defer the full definition of the triangle tilt to Definition 2.12 in Section 2.2. We shall show that in a number of different regimes, the triangle tilt is the best possible tilt, in an asymptotic sense. In this sense, the class of exponential random graphs is sufficiently rich to ensure the existence of an optimal triangle tilt even for a subset of the replica breaking phase.

To understand why the class of exponential random graphs is the right class to consider, we make a digression to mention the connection between exponential random graphs and the conditioned Erdős-Rényi graphs. Exponential random graphs have been studied in [1, 5, 16, 17]. The “classical” exponential random graphs with $\alpha = 1$ and its connection to conditioned Erdős-Rényi graphs was initially observed by Chatterjee and Dey [7] when proving the large deviations principle (1.2), and it was further developed by Lubetzky and Zhao [16] for $\alpha > 0$. An interesting observation in [16], for the case when (p, t) belongs to the replica symmetric phase, is the connection between the free energy of the Gibbs measure and the derivative of the rate function. This connection leads to the following duality relationship between certain parameters (h, β, α) of the Gibbs measure and the parameters (p, t) of the conditioned Erdős-Rényi graph: for (p, t) that is replica symmetric and $\alpha \in [2/3, 1]$, the typical exponential random graph resembles the conditioned Erdős-Rényi graph if $h = \log \frac{p}{1-p}$ and the free energy of the Gibbs measure $\mathbb{Q}_n^{h,\beta,\alpha}$, expressed in a variational formulation

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \sum_X e^{H(X)} = \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - I_p(u) \right]$$

where $I_p(t)$ is the rate function at (1.3), is maximized at t .

One of our main results, Theorem 2.5, extends this duality into the replica breaking phase, and generalizes the way of characterizing when an exponential random graph resembles an conditioned Erdős-Rényi graph. The gist of Theorem 2.5, and its immediate consequence is described, heuristically, as follows:

Fix $p \in (0, 1)$ and $t \in [p, 1]$, and let $h_p = \log \frac{p}{1-p}$. Suppose there exists $\beta > 0$ and $\alpha \in [0, 1]$ such that

$$t = \arg \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \phi_p(u) \right], \quad (1.4)$$

where the function $\phi_p(u)$ is a rate function (see (2.24)). Then the exponential random graph $\mathcal{G}_n^{h_p,\beta,\alpha}$ and the Erdős-Rényi graph $\mathcal{G}_{n,p}$ conditioned on $\{T(\mathcal{G}_{n,p}) \geq \binom{n}{3} t^3\}$ are asymptotically indistinguishable.

Thus, Theorem 2.5 provides a way to characterize the asymptotic behaviour of the conditioned Erdős-Rényi graph by that of an exponential random graph.

Apart from its independent interest, Theorem 2.5 and the variational form (1.4) is the basis for choosing the parameters for the Gibbs measure that defines the triangle tilt. In essence, the triangle tilt can be defined when there exists (h_p, β, α) for which (1.4) holds,

which is the case for the replica symmetric phase and at least a nontrivial subset of the replica breaking phase.

Returning to the question of the efficiency of an importance sampling scheme, one measure of the efficiency is through the magnitude of the variance of the importance sampling estimator. In the presence of a large deviation principle, we appeal to the notion of *asymptotic optimality*, which is the property that the importance sampling estimator has the smallest attainable variance, as $n \rightarrow \infty$. (See Definition 2.3.)

Our main results pertain to the asymptotic optimality or non-optimality of certain importance sampling estimators. In Proposition 3.1 we prove a necessary condition for asymptotic optimality when the tilt is based on an exponential random graph: the exponential random graph must be asymptotically indistinguishable from the conditioned Erdős-Rényi graph. In particular, if (p, t) belongs to the replica symmetric phase, then the necessary condition is that the exponential random graph is indistinguishable from $\mathcal{G}_{n,t}$. On the other hand, Proposition 3.2 shows that this is not a sufficient condition for asymptotic optimality: there is a subregime of the replica symmetric phase for which the edge tilt produces a suboptimal estimator. It is interesting to note that although the LDP suggests that $\mathcal{G}_{n,t}$ is the typical behaviour of the conditioned ER graph in the replica symmetric phase, directly using $\mathcal{G}_{n,t}$ as the importance sampling tilt does not necessarily give an optimal estimator. Instead, we must be careful to use a tilt that not only is indistinguishable from the conditioned Erdős-Rényi graph, but also gives an asymptotically optimal estimator. It turns out that the triangle tilts are the appropriate tilts to use, and this fact is the statement of our main optimality result, which we state here.

Theorem 1.1. *Given (p, t) , denote $h_p = \log \frac{p}{1-p}$. Suppose there exists a triangle tilt $\mathbb{Q}_n^{h_p, \beta, \alpha}$ with parameter $\alpha > 0$ corresponding to (p, t) , as defined in Definition 2.12. Then the importance sampling estimator based on the tilted measure $\mathbb{Q}_n^{h_p, \beta, \alpha}$ is asymptotically optimal.*

Organization of the paper: We start by giving precise definitions of the various constructs arising in our study in Section 2. This culminates in Theorem 2.5 that characterizes the limiting free energy of the exponential random graph model. The rest of Section 2 is devoted to drawing a connection between the exponential random graph and Erdős-Rényi random graph conditioned on an atypical number of triangles, leading to the derivation of the triangle tilts. Section 3 discusses and proves our main results on asymptotic optimality or non-optimality of the importance sampling estimators. In Section 4, we carry out numerical simulations on moderate size networks using the various proposed tilts to illustrate and compare the viability of the importance sampling schemes.

Acknowledgement This work was funded in part through the 2011-2012 SAMSI Program on Uncertainty Quantification, in which each of the authors participated. JN was partially supported by grant NSF-DMS 1007572. SB was partially supported by grant NSF-DMS 1105581.

2. LARGE DEVIATIONS, IMPORTANCE SAMPLING AND EXPONENTIAL RANDOM GRAPHS

A simple graph X on n vertices can be represented as an element of the space $\Omega_n = \{0, 1\}^{\binom{n}{2}}$. A graph $X \in \Omega_n$ will be denoted by $X = (X_{ij})_{1 \leq i < j \leq n}$ with the entry X_{ij} indicating the presence or absence of an edge between vertices i and j . For a given edge probability $p \in [0, 1]$, an Erdős-Rényi random graph $\mathcal{G}_{n,p}$ is a graph on n vertices such that

any edge is independently connected with probability p . We shall use $\mathbb{P}_{n,p}$ to represent the probability measure on Ω_n induced by the Erdős-Rényi graph $\mathcal{G}_{n,p}$. The probability of a fixed graph X under the measure $\mathbb{P}_{n,p}$ can be explicitly computed as

$$\mathbb{P}(G_{n,p} = X) = \mathbb{P}_{n,p}(X) = \prod_{i < j} p^{X_{ij}} (1-p)^{1-X_{ij}} = (1-p)^{\binom{n}{2}} e^{h_p E(X)} \quad (2.1)$$

where $h_p := \log \frac{p}{1-p}$, and $E(X) := \sum_{i < j} X_{ij}$ is the number of edges in X . Let $T(X)$ denote the number of triangles in graph X :

$$T(X) = \sum_{1 \leq i < j < k \leq n} X_{ij} X_{jk} X_{ik}.$$

Also let the event $W_{n,t} = \{X \in \Omega_n \mid T(X) \geq \binom{n}{3} t^3\}$ denote the upper tails of triangle counts.

Importance sampling. If $\{X^k\}_{k=1}^\infty \subset \Omega_n$ is a sequence of Erdős-Rényi random graphs generated independently from $\mathbb{P}_{n,p}$, then for any integer $K \geq 1$,

$$M_K = \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{W_{n,t}}(X^k)$$

is an unbiased estimate of μ_n . By the law of large numbers, $M_K \rightarrow \mu_n$ with probability one as $K \rightarrow \infty$. Although this estimate of μ_n is very simple, the relative error is

$$\frac{\sqrt{\text{Var}(M_K)}}{\mathbb{E}(M_K)} = \frac{\sqrt{\mu_n - (\mu_n)^2}}{\mu_n \sqrt{K}},$$

which scales like $(K\mu_n)^{-1/2}$ as $\mu_n \rightarrow 0$. Hence the relative error may be very large in the large deviation regime where $\mu_n \ll 1$, unless we have at least $K \sim O(\mu_n^{-1})$ samples. Therefore, it is desirable to devise an estimate of μ_n which, compared to this simple Monte Carlo estimate, attains the same accuracy with fewer number of samples or lower computational cost.

Importance sampling is a Monte Carlo algorithm based on a change of measure. Suppose that $\mathbb{P}_{n,p}$ is absolutely continuous with respect to another measure \mathbb{Q} on Ω_n with

$$\frac{d\mathbb{P}_{n,p}}{d\mathbb{Q}} = Y^{-1} : \Omega_n \rightarrow \mathbb{R}.$$

Then we have

$$\mu_n = \mathbb{E}[M_K] = \mathbb{E} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{1}_{W_{n,t}}(X^k) \right] = \mathbb{E}_{\mathbb{Q}} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{1}_{W_{n,t}}(\tilde{X}^k) Y^{-1}(\tilde{X}^k) \right] \quad (2.2)$$

where $\mathbb{E}_{\mathbb{Q}}$ denotes expectation with respect to \mathbb{Q} , and we now use $\{\tilde{X}^k\}_{k=1}^\infty$ to denote a set of random graphs sampled independently from the new measure \mathbb{Q} . If we define

$$\tilde{M}_K = \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{W_{n,t}}(\tilde{X}^k) Y^{-1}(\tilde{X}^k), \quad (2.3)$$

then \tilde{M}_K is also an unbiased estimate of μ_n , and the relative error is now:

$$\frac{\sqrt{\text{Var}_{\mathbb{Q}}(M_K)}}{\mathbb{E}_{\mathbb{Q}}(M_K)} = \frac{\sqrt{\mathbb{E}_{\mathbb{Q}}[(\mathbf{1}_{W_{n,t}}(X) Y^{-1})^2] - (\mu_n)^2}}{\mu_n \sqrt{K}}, \quad (2.4)$$

Formally this is optimized by the choice $Y = (\mu_n)^{-1} \mathbf{1}_{W_{n,t}}(X)$, in which case the relative error is zero. Such a choice for \mathbb{Q} is not feasible, however, since normalizing Y would require a priori knowledge of $\mu_n = \mathbb{P}_{n,p}(W_{n,t})$. Intuitively, we should choose the tilted measure \mathbb{Q} so that $\tilde{X}_k \in W_{n,t}$ occurs with high probability under \mathbb{Q} .

We will refer to Y^{-1} as the importance sampling weights, and \mathbb{Q} as the *tilted measure*, or *tilt*. If \mathbb{Q} arises naturally as the measure induced by a random graph \mathcal{G}_n , we will also refer to \mathcal{G}_n as the *tilt*.

Asymptotic optimality and large deviations. In view of (2.3), one way of comparing the efficiency of importance sampling estimators is to look at which estimator has the smaller second moment. When the family of measures $\mathbb{P}_{n,p}$ possesses a large deviations principle, a notion of asymptotic optimality (or asymptotic efficiency) of the estimator \tilde{M}_K can be defined with the interpretation that the second moment of the estimator is the smallest possible in the asymptotic sense, as afforded by the large deviation principle ([4] and see also Definition 2.3). Thus, before defining asymptotic optimality, we first proceed with a description of the large deviations principle for Erdős-Rényi random graphs.

In the context of Erdős-Rényi random graphs, Chatterjee and Varadhan [8] have proved a general large deviation principle which is based on the theory of dense graph limits developed by [3]. In this framework, a random graph is represented as a function $X(x, y) \in \widetilde{\mathcal{W}}$, where $\widetilde{\mathcal{W}}$ is the set of all measurable functions $f : [0, 1]^2 \rightarrow [0, 1]$ satisfying $f(x, y) = f(y, x)$. Specifically, a finite simple graph X on n vertices is represented by the function, or *graphon*,

$$X(x, y) = \sum_{\substack{i, j=1 \\ i \neq j}}^n X_{ij} \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n}) \times [\frac{j-1}{n}, \frac{j}{n})}(x, y) \in \widetilde{\mathcal{W}}. \quad (2.5)$$

Here we treat (X_{ij}) as a symmetric matrix with entries in $\{0, 1\}$ and $X_{ii} = 0$ for all i . In general, for a function $f \in \widetilde{\mathcal{W}}$, $f(x, y)$ can be interpreted as the probability of having an edge between vertices x and y . Then, we define the quotient space \mathcal{W} under the equivalence relation defined by $f \sim g$ if $f(x, y) = g(\sigma x, \sigma y)$ for some measure preserving bijection $\sigma : [0, 1] \rightarrow [0, 1]$. Intuitively, an equivalence class contains graphons that are equal after a relabelling of vertices. (See, e.g., [3, 8] for further exploration and properties of the quotient space.)

By identifying a finite graph X with its graphon representation, we can consider the probability measure $\mathbb{P}_{n,p}$ as a measure induced on \mathcal{W} supported on the subset of graphons of finite graphs. For $f \in \mathcal{W}$, denote

$$\mathcal{E}(f) = \int_0^1 \int_0^1 f(x, y) dx dy \quad (2.6)$$

and

$$\mathcal{T}(f) = \int_0^1 \int_0^1 \int_0^1 f(x, y) f(y, z) f(x, z) dx dy dz. \quad (2.7)$$

We see that $E(X) = \frac{n^2}{2} \mathcal{E}(X)$ and $T(X) = \frac{n^3}{6} \mathcal{T}(X)$, so that \mathcal{E} and \mathcal{T} represent edge and triangle densities of the graph X , respectively. Then, rather than considering the event $W_{n,t}$, we shall equivalently consider the upper tails of triangle densities,

$$\mathcal{W}_t := \{f \in \mathcal{W} \mid \mathcal{T}(f) \geq t^3\}.$$

The large deviation principle of Chatterjee and Varadhan [8] implies for any $p \in (0, 1)$ and $t \in [p, 1]$,

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\mathcal{T}(\mathcal{G}_{n,p}) \geq t^3) = -\phi(p, t) \quad (2.8)$$

where $\phi(p, t)$ is the large deviation decay rate given by a variational form,

$$\phi(p, t) = \inf \{ \mathcal{I}_p(f) \mid f \in \mathcal{W}, \mathcal{T}(f) \geq t^3 \} = \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]. \quad (2.9)$$

Here,

$$\mathcal{I}_p(f) := \int_0^1 \int_0^1 I_p(f(x, y)) dx dy \quad (2.10)$$

is the large deviation rate function, where $I_p : [0, 1] \rightarrow \mathbb{R}$ is defined at (1.3). A further important consequence of the large deviation principle concerns the typical behaviour of the conditioned probability measure

$$\mathbb{P}_{n,p}(X | \mathcal{W}_t) = \mathbb{P}_{n,p}(X) \mathbf{1}_{\mathcal{W}_t}(X) \mu_n^{-1}.$$

When we refer to $\mathcal{G}_{n,p}$ conditioned on the event $\{\mathcal{T}(f) \geq t^3\}$, we mean the random graph whose law is given by this conditioned probability measure.

Lemma 2.1. ([8, Theorem 3.1], Lemma A.1) *Let $\mathcal{F}^* \subset \mathcal{W}$ be the set of graphs that optimize the variational form in (2.9). Then the Erdős-Rényi graph $\mathcal{G}_{n,p}$ conditioned on $\{\mathcal{T}(f) \geq t^3\}$ is asymptotically indistinguishable from the minimal set \mathcal{F}^* .*

The term “asymptotically indistinguishable” in Lemma 2.1 roughly means that the graphon representation of the graph converges in probability, under the *cut distance* metric, to the constant function u^* at an exponential rate as $n \rightarrow \infty$. Intuitively, this means that the typical conditioned Erdős-Rényi graph resembles some graph $f^* \in \mathcal{F}^*$ for large n . In order to give a more precise definition of asymptotic indistinguishability, we first recall the cut distance metric δ_\square , defined for $f, g \in \mathcal{W}$ by

$$\delta_\square(f, g) = \inf_{\sigma} \sup_{S, T \subset [0, 1]} \left| \int_{S \times T} (f(\sigma x, \sigma y) - g(x, y)) dx dy \right|,$$

where the infimum is taken over all measure-preserving bijections $\sigma : [0, 1] \rightarrow [0, 1]$. For $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{W}$,

$$\delta_\square(\mathcal{F}_1, \mathcal{F}_2) = \inf_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} \delta_\square(f_1, f_2).$$

It is known by [14] that $(\mathcal{W}, \delta_\square)$ is a compact metric space.

We say that a random graph \mathcal{G}_n on n vertices is asymptotically indistinguishable from a subset $\mathcal{F} \subset \mathcal{W}$ if for any $\epsilon_1 > 0$ there is $\epsilon_2 > 0$ such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\delta_\square(\mathcal{G}_n, \mathcal{F}) > \epsilon_1) < -\epsilon_2. \quad (2.11)$$

Further, we say that \mathcal{G}_n is asymptotically indistinguishable from the *minimal* set $\mathcal{F} \subset \mathcal{W}$ if \mathcal{F} is the smallest closed subset of \mathcal{W} that \mathcal{G}_n is asymptotically indistinguishable from. Clearly, if \mathcal{G}_n is asymptotically indistinguishable from a singleton set \mathcal{F} , then \mathcal{F} is, trivially, minimal. Finally, we say two random graphs $\mathcal{G}_n^1, \mathcal{G}_n^2$ are asymptotically indistinguishable if they are each asymptotically indistinguishable from the same minimal set $\mathcal{F} \subset \mathcal{W}$. Intuitively, this means that the random behaviour, or the typical graphs, of \mathcal{G}_n^1 resembles that of \mathcal{G}_n^2 for large n . (See [5] and [8] for a wide-ranging exploration of this metric in the context of describing limits of dense random graph sequences.)

Using this terminology, we observe that an Erdős-Rényi graph $\mathcal{G}_{n,u}$ is asymptotically indistinguishable from the singleton set containing the constant function $f^* \equiv u$. A special notion about whether the conditioned Erdős-Rényi graph is again an Erdős-Rényi graph leads to the following definition.

Definition 2.2. The *replica symmetric phase* is the regime of parameters (p, t) for which the large deviations rate satisfies

$$\inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)] = I_p(t), \quad (2.12)$$

and the infimum is uniquely attained at the constant function t .

The *replica breaking phase* is the regime of parameters (p, t) that are not in the replica symmetric phase. \blacksquare

Hence, the notion of replica symmetry is a property of the rare event problem, where, conditioned on the event $\{\mathcal{T}(f) \geq t^3\}$, the Erdős-Rényi graph $\mathcal{G}_{n,p}$ behaves like another Erdős-Rényi graph $\mathcal{G}_{n,t}$ with the higher edge density t , for large n . In contrast, the conditioned graph in the replica breaking phase is not any Erdős-Rényi graph, and has been conjectured to exhibit a clique-like structure with lesser than t edge density. The term “replica symmetric phase” is borrowed from [8], which in turn was inspired by the statistical physics literature. However, we remark that this term has been used by different authors to refer to other instances of graphs behaving like an Erdős-Rényi graph.

The large deviations principle gives us an estimate of the relative error in the estimate \tilde{M}_K . For any fixed K , it is clear from (2.4) that minimizing the relative error is equivalent to minimizing the second moment $\mathbb{E}_{\mathbb{Q}_n}[(\mathbf{1}_{\mathcal{W}_t} Y^{-1})^2]$. By Jensen’s inequality, we have the following asymptotic lower bound:

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}_n}[(\mathbf{1}_{\mathcal{W}_t} Y^{-1})^2] \geq -2 \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)] = -2\phi(p, t). \quad (2.13)$$

This leads to the definition of asymptotic optimality.

Definition 2.3. A family of tilted measures \mathbb{Q}_n on \mathcal{W} is said to be *asymptotically optimal* if

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}_n}[(\mathbf{1}_{\mathcal{W}_t} Y^{-1})^2] = -2 \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)].$$

In contrast, the second moment of each term in the simple Monte Carlo method satisfies

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{P}_n}[\mathbf{1}_{\mathcal{W}_t}^2] = - \inf_{f \in \mathcal{W}_t} \mathcal{I}_p(f) = -\phi(p, t) > -2\phi(p, t).$$

Thus, the simple Monte Carlo method is not asymptotically optimal. Observe that Jensen’s inequality for conditional expectation implies

$$\begin{aligned} \mathbb{Q}_n(\mathcal{W}_t)^{-1} &= \mathbb{P}_n(\mathcal{W}_t)^{-1} \left(\frac{\mathbb{E}_{\mathbb{P}_n}(\mathbf{1}_{\mathcal{W}_t} Y)}{\mathbb{P}_n(\mathcal{W}_t)} \right)^{-1} \\ &\leq \mathbb{P}_n(\mathcal{W}_t)^{-2} \mathbb{E}_{\mathbb{P}_n}(\mathbf{1}_{\mathcal{W}_t} Y^{-1}) = \mathbb{P}_n(\mathcal{W}_t)^{-2} \mathbb{E}_{\mathbb{Q}_n}(\mathbf{1}_{\mathcal{W}_t} Y^{-2}). \end{aligned} \quad (2.14)$$

So, if \mathbb{Q}_n is asymptotically optimal, we must have

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{Q}_n(\mathcal{W}_t) \geq \liminf_{n \rightarrow \infty} \frac{2}{n^2} \log \mathbb{P}_n(\mathcal{W}_t) + \liminf_{n \rightarrow \infty} \frac{-1}{n^2} \log \mathbb{E}_{\mathbb{Q}_n}(\mathbf{1}_{\mathcal{W}_t} Y^{-2}) = 0, \quad (2.15)$$

which is consistent with the intuition that a good choice of \mathbb{Q}_n should put $\tilde{X}_k \in \mathcal{W}_t$ with high probability.

2.1. Asymptotic behavior of exponential random graphs. To find “good” importance sampling tilted measures, we focus on the class of exponential random graphs. The exponential random graph is a random graph on n vertices defined by the Gibbs measure

$$\mathbb{Q}(X) = \mathbb{Q}_n^{h,\beta,\alpha}(X) \propto e^{n^2 \mathcal{H}(X)} \quad (2.16)$$

on Ω_n , where for given $h \in \mathbb{R}$, $\beta \in \mathbb{R}_+$, $\alpha > 0$, the Hamiltonian is

$$\mathcal{H}(X) = \frac{h}{2} \mathcal{E}(X) + \frac{\beta}{6} \mathcal{T}(X)^\alpha. \quad (2.17)$$

We will use $\psi_n = \psi_n^{h,\beta,\alpha}$ to denote the log of the normalizing constant (free energy)

$$\psi_n = \psi_n^{h,\beta,\alpha} = \frac{1}{n^2} \log \sum_{X \in \Omega_n} e^{n^2 \mathcal{H}(X)},$$

so that $\mathbb{Q}_n^{h,\beta,\alpha} = \exp(n^2(\mathcal{H}(X) - \psi_n))$. We denote by $\mathcal{G}_n^{h,\beta,\alpha}$ the exponential random graph defined by the Gibbs measure (2.16). The case where $\alpha = 1$ is the “classical” exponential random graph model that has an enormous literature in the social sciences, see e.g. [18, 19] and the references therein and rigorously studied in a number of recent papers, see e.g. [1, 5, 16, 17, 21, 22]. In this case, the Hamiltonian can be rewritten as $n^2 \mathcal{H}(X) = hE(X) + \frac{\beta}{n} T(X)$. We will drop the superscripts in $\psi_n^{h,\beta}$, $\mathbb{Q}_n^{h,\beta}$ when $\alpha = 1$. The generalization to the exponential random graph with the parameter α was first proposed in [16].

Observe that the Erdős-Rényi random graph is a special case of the exponential random graph: if $\beta = 0$ and $h = h_p$ with h_p defined by (2.1), then $\mathbb{Q}_n^{h_p,0,\alpha} = \mathbb{P}_{n,p}$ for any $\alpha > 0$ and the edges are independent with probability p . On the other hand, choosing $\beta > 0$ introduces a non-trivial dependence between the edges. By adjusting the parameters (h, β, α) , the Gibbs measure $\mathbb{Q}_n^{h,\beta,\alpha}$ can be adjusted to favor edges and triangles to varying degree.

The asymptotic behavior of the exponential random graph measures $\mathbb{Q}_n^{h,\beta,\alpha}$ and the free energy $\psi_n^{h,\beta,\alpha}$ is partially characterized by the following result of Chatterjee and Diaconis [5] and Lubetzky and Zhao [16]. In what follows, we will make use of the functions

$$I(u) = \frac{1}{2} u \log u + \frac{1}{2} (1-u) \log(1-u) \quad (2.18)$$

on $u \in [0, 1]$ and, for $f \in \mathcal{W}$,

$$\mathcal{I}(f) := \int_0^1 \int_0^1 I(f(x, y)) \, dx \, dy. \quad (2.19)$$

Theorem 2.4. (a) [5, Theorems 4.1, 4.2] *For the classical exponential random graph with $\alpha = 1$, the free energy satisfies*

$$\lim_{n \rightarrow \infty} \psi_n^{h,\beta} = \sup_{0 \leq u \leq 1} \left[\frac{h}{2} u + \frac{\beta}{6} u^3 - I(u) \right]. \quad (2.20)$$

If the supremum in (2.20) is attained at a unique point $u^ \in [0, 1]$, then the exponential random graph $\mathcal{G}_n^{h,\beta}$ is asymptotically indistinguishable from the Erdős-Rényi graph \mathcal{G}_{n,u^*} .*

(b) [16, Theorems 1.3, 4.3] *For the exponential random graph with parameter $\alpha \in [2/3, 1]$, the free energy satisfies*

$$\lim_{n \rightarrow \infty} \psi_n^{h,\beta,\alpha} = \sup_{0 \leq u \leq 1} \left[\frac{h}{2} u + \frac{\beta}{6} u^{3\alpha} - I(u) \right]. \quad (2.21)$$

If the supremum in (2.21) is attained at a unique point $u^* \in [0, 1]$, then the exponential random graph $\mathcal{G}_n^{h, \beta, \alpha}$ is asymptotically indistinguishable from the Erdős-Rényi graph \mathcal{G}_{n, u^*} .

Our main result in this section, stated next, is the generalization of the variational formulation for the free energy of the Gibbs measure of any exponential random graph. The consequence of this result leads to the connection between the exponential random graph and the conditioned Erdős-Rényi graph. Before stating the result we will need some extra notation. Extend the Hamiltonian defined in (2.17) to the space of graphons in the natural way

$$\mathcal{H}(f) := \frac{h}{2}\mathcal{E}(f) + \frac{\beta}{6}\mathcal{T}(f) \quad (2.22)$$

where recall the definitions for the density of edges and triangles for graphons defined respectively in (2.6) and (2.7). For fixed $q \in (0, 1)$ recall the functions $\mathcal{I}_q(f)$ from (2.10) and the function $\mathcal{I}(f)$ from (2.19).

Theorem 2.5. *Given any Gibbs measure parameters $(h, \beta, \alpha) \in \mathbb{R} \times \mathbb{R}^+ \times (0, 1]$, assume wlog that $h = h_q = \log \frac{q}{1-q}$ for some $q \in (0, 1)$. For $u \in [0, 1]$, denote $\partial\mathcal{W}_u := \{f \in \mathcal{W} \mid \mathcal{T}(f) = u^3\}$ and let $\mathcal{F}_u^* \subset \mathcal{W}$ be the set of minimizers of $\inf_{f \in \partial\mathcal{W}_u} [\mathcal{I}_q(f)]$. Then for the exponential random graph $\mathcal{G}_n^{h_q, \beta, \alpha}$, the free energy satisfies*

$$\begin{aligned} \lim_{n \rightarrow \infty} \psi_n^{h_q, \beta, \alpha} &= \sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] \\ &= \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \phi_q(u) - \frac{1}{2} \log(1 - q) \right] \end{aligned} \quad (2.23)$$

where

$$\phi_q(u) = \inf_{f \in \partial\mathcal{W}_u} [\mathcal{I}_q(f)]. \quad (2.24)$$

The supremum, $\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)]$, is attained exactly on the set $\mathcal{F}_{v^*}^*$, where v^* maximizes the RHS of (2.23).

Further, if (q, v^*) belongs to the replica symmetric phase, then the supremum, $\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)]$, is attained uniquely by the constant function $f_{v^*} \equiv v^*$, and

$$\lim_{n \rightarrow \infty} \psi_n^{h_q, \beta, \alpha} = V(v^*) = \sup_{0 \leq u \leq 1} [V(u)] \quad (2.25)$$

where

$$V(u) = V(u; h_q, \beta, \alpha) = \frac{h_q}{2}u + \frac{\beta}{6}u^{3\alpha} - \mathcal{I}(u).$$

Proof. The first equality in (2.23) follows from Thm 3.1 in [5]. To show the second equality, suppose $f \in \partial\mathcal{W}_u$, for $u \in (0, 1)$.

$$\begin{aligned} \mathcal{H}(f) - \mathcal{I}(f) &= \frac{h_q}{2}\mathcal{E}(f) + \frac{\beta}{6}u^{3\alpha} - \mathcal{I}(f) \\ &= \frac{\beta}{6}u^{3\alpha} - \mathcal{I}_q(f) - \frac{1}{2}\log(1 - q) \\ &\leq \frac{\beta}{6}u^{3\alpha} - \inf_{f \in \partial\mathcal{W}_u} [\mathcal{I}_q(f)] - \frac{1}{2}\log(1 - q) \end{aligned} \quad (2.26)$$

This implies that

$$\sup_{f \in \partial\mathcal{W}_u} [\mathcal{H}(f) - \mathcal{I}(f)] \leq \frac{\beta}{6}u^{3\alpha} - \inf_{f \in \partial\mathcal{W}_u} [\mathcal{I}_q(f)] - \frac{1}{2}\log(1 - q),$$

and the supremum $\sup_{f \in \partial \mathcal{W}_u} [\mathcal{H}(f) - \mathcal{I}(f)]$ is attained on the same set of functions $\mathcal{F}_u^* \subset \mathcal{W}$ that optimize $\inf_{f \in \partial \mathcal{W}_u} [\mathcal{I}_q(f)]$. Then,

$$\begin{aligned} \sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] &= \sup_{0 \leq u \leq 1} \sup_{f \in \partial \mathcal{W}_u} [\mathcal{H}(f) - \mathcal{I}(f)] \\ &\leq \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \inf_{f \in \partial \mathcal{W}_u} [\mathcal{I}_q(f)] - \frac{1}{2} \log(1 - q) \right] \end{aligned}$$

For each $u \in (0, 1)$, let $f_u \in \mathcal{F}_u^*$. Let v^* maximize the RHS of (2.23). Then

$$\begin{aligned} v^* &:= \arg \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \inf_{f \in \partial \mathcal{W}_u} [\mathcal{I}_q(f)] - \frac{1}{2} \log(1 - q) \right] \\ &= \arg \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \mathcal{I}_q(f_u) - \frac{1}{2} \log(1 - q) \right]. \end{aligned}$$

It follows that

$$\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] = \frac{\beta}{6} (v^*)^{3\alpha} - \mathcal{I}_q(f_{v^*}) - \frac{1}{2} \log(1 - q)$$

and moreover, the supremum $\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)]$ is attained by any $f_{v^*} \in \mathcal{F}_{v^*}^*$. This concludes the proof of (2.23).

Now suppose (q, v^*) belongs to the replica symmetric phase. This implies that the constant function v^* is the unique minimizer of the LDP rate function $\inf_{f \in \mathcal{W}_{v^*}} [\mathcal{I}_q(f)]$, and by Theorem 4.2(iii) in [8],

$$\mathcal{I}_q(v^*) = \inf_{f \in \mathcal{W}_{v^*}} [\mathcal{I}_q(f)] = \inf_{f \in \partial \mathcal{W}_{v^*}} [\mathcal{I}_q(f)].$$

Since $I_q(u) = I(u) - \frac{h_q}{2}u - \frac{1}{2} \log(1 - u)$, we have that the free energy is

$$\begin{aligned} \lim_{n \rightarrow \infty} \psi_n^{h_q, \beta, \alpha} &= \frac{\beta}{6} (v^*)^{3\alpha} - \inf_{f \in \partial \mathcal{W}_{v^*}} [\mathcal{I}_q(f)] - \frac{1}{2} \log(1 - q) \\ &= \frac{\beta}{6} (v^*)^{3\alpha} - I_q(v^*) - \frac{1}{2} \log(1 - q) \\ &= \frac{h}{2} v^* + \frac{\beta}{6} (v^*)^{3\alpha} - I(v^*) = V(v^*). \end{aligned}$$

Moreover, we claim that $V(v^*) = \sup_{0 \leq u \leq 1} [V(u)]$. To see this, notice that if the optimizer of $\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)]$ is a constant function, then it suffices to consider the supremum only over constant functions. But the only constant function in $\partial \mathcal{W}_u$ is the function u . So

$$\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] = \sup_{f \in \mathcal{C}} [\mathcal{H}(f) - \mathcal{I}(f)] = \sup_{0 \leq u \leq 1} [\mathcal{H}(u) - I(u)] = \sup_{0 \leq u \leq 1} [V(u)]$$

where $\mathcal{C} \subset \mathcal{W}$ is the set of constant functions. The proof is complete. \blacksquare

Remark 2.6. Recalling the LDP rate $\phi(q, u)$ defined in (2.9), [8, Theorem 4.2(iii)] showed that for $u \geq q$

$$\phi(q, u) = \inf_{f \in \mathcal{W}_u} [\mathcal{I}_q(f)] = \inf_{f \in \partial \mathcal{W}_u} [\mathcal{I}_q(f)] = \phi_q(u),$$

and the set of minimizers that attain the rate $\phi(q, u)$ is exactly \mathcal{F}_u^* . So, for any $u \geq q$, the Erdős-Rényi graph $\mathcal{G}_{n,q}$ conditioned on the event $\{\mathcal{T}(f) \geq u^3\}$ is asymptotically indistinguishable from the minimal set \mathcal{F}_u^* , by Lemma 2.1. On the other hand, since the set $\mathcal{F}_{v^*}^*$ attains the supremum, $\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)]$, in Theorem 2.5, it follows from [5,

Theorem 3.2] and Lemma A.1 that the exponential random graph $\mathcal{G}_n^{h_q, \beta, \alpha}$ is asymptotically indistinguishable from the minimal set $\mathcal{F}_{v^*}^*$. We have the following corollary.

Corollary 2.7. *Let the parameters (h_q, β, α) , (q, v^*) be as in Theorem 2.5 and Eqn (2.23). Suppose $v^* \geq q$. Then $\mathcal{G}_n^{h_q, \beta, \alpha}$ is asymptotically indistinguishable from the conditioned Erdős-Rényi graph, $\mathcal{G}_{n,q}$ conditioned on the event $\{\mathcal{T}(f) \geq (v^*)^3\}$.*

In particular, if (q, v^) belongs to the replica symmetric phase, then $\mathcal{G}_n^{h_q, \beta, \alpha}$ is asymptotically indistinguishable from the Erdős-Rényi graph \mathcal{G}_{n, v^*} .*

The mean behaviour of the triangle density of an exponential random graph $\mathcal{G}_n^{h_q, \beta, \alpha}$ can be deduced from the variational formulation in (2.23), and in special instances, so can the mean behaviour of the edge density. This is shown in the next proposition.

Proposition 2.8. *Given (h_q, β, α) as in Theorem 2.5, if the supremum in (2.23) is attained at a unique point $v^* \in [0, 1]$, then*

$$\lim_{n \rightarrow \infty} \mathbb{E} |\mathcal{T}(\mathcal{G}_n^{h_q, \beta, \alpha}) - (v^*)^3| = 0. \quad (2.27)$$

Further, if (q, v^) belongs to the replica symmetric phase, then*

$$\lim_{n \rightarrow \infty} \mathbb{E} |\mathcal{E}(\mathcal{G}_n^{h_q, \beta, \alpha}) - v^*| = 0. \quad (2.28)$$

Proof. This follows from [5, Theorem 4.2] and the Lipschitz continuity of the mappings $f \mapsto \mathcal{T}(f)$ and $f \mapsto \mathcal{E}(f)$ under the cut distance metric δ_\square [3, Theorem 3.7]. The proof is left to the appendix. \blacksquare

2.2. Triangle and edge tilts. In this section, we use the variational form of the free energy, (2.23), to construct the triangle tilts for the importance sampling scheme (see Definition 2.12). In order to define the triangle tilts, and in view of Theorem 2.5, we must characterize the (p, t) regime where there exists a Gibbs measure $\mathbb{Q}_n^{h_p, \beta, \alpha}$ satisfying

$$t = \arg \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \inf_{f \in \partial \mathcal{T}_u} [\mathcal{I}_p(f)] - \frac{1}{2} \log(1 - p) \right] = \arg \sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \phi_p(u) \right] \quad (2.29)$$

where $\phi_p(u) = \inf_{f \in \partial \mathcal{W}_u} [\mathcal{I}_p(f)]$. Since $\mathcal{I}_p(f)$ is the rate function, it is known that $\phi_p(p) = 0$, and $\phi_p(u)$ is continuous and strictly increasing on $[p, 1]$ (Theorem 4.3 in [8]).

If $\phi_p(u)$ is differentiable everywhere, then the extremal points u^* of the function $\tilde{V}(u) := \frac{\beta}{6} u^{3\alpha} - \phi_p(u)$ satisfies

$$\tilde{V}'(u^*) = \frac{\beta\alpha}{2} (u^*)^{3\alpha-1} - \phi_p'(u^*) = 0,$$

Then for (2.29) to hold, β must necessarily be given by

$$\beta^* = \frac{2\phi_p'(t)}{\alpha t^{3\alpha-1}}. \quad (2.30)$$

The next lemma shows that, regardless of the differentiability of $\phi_p(t)$, provided a certain *minorant condition* holds, we can find a β and a sufficiently small α so that (2.29) holds, and consequently that the exponential graph is asymptotically indistinguishable from the conditioned Erdős-Rényi graph.

We shall say that (p, t) satisfies the *minorant condition with parameter α* if $(t^{3\alpha}, \phi_p(t))$ lies on the convex minorant of the function $x \mapsto \phi_p(x^{1/3\alpha})$. If $(t^{3\alpha}, \phi_p(t))$ lies on the convex minorant of $x \mapsto \phi_p(x^{1/3\alpha})$, then subdifferential(s) of the convex minorant of $x \mapsto \phi_p(x^{1/3\alpha})$ always exist and are positive. Recall that the subdifferentials of a convex function $f(x)$

at a point x are the slopes of any line lying below $f(x)$ that is tangent to f at x . The set of subdifferentials of a convex function is non-empty; if the function is differentiable at x , then the set of subdifferentials contains exactly one point, the derivative $f'(x)$.

Lemma 2.9. *Suppose (p, t) satisfies the minorant condition for $\alpha > 0$ sufficiently small. Let $\frac{\beta}{6}$ be any subdifferential of the convex minorant of $x \mapsto \phi_p(x^{1/3\alpha})$ at the point $t^{3\alpha}$. Then $\sup_{0 \leq u \leq 1} [\frac{\beta}{6}u^{3\alpha} - \phi_p(u)]$ is maximized at t . Moreover, if $\phi_p(u)$ is differentiable at t , then $\beta = \beta^*$, as defined in (2.30).*

Proof. The proof follows a similar technique to [16]. Using the rescaling $u \mapsto x^{1/3\alpha}$, the variational form $\sup_{0 \leq u \leq 1} [\frac{\beta}{6}u^{3\alpha} - \phi_p(u)]$ can be rewritten as

$$\sup_{0 \leq x \leq 1} [\frac{\beta}{6}x - \phi_p(x^{1/3\alpha})].$$

Let $\hat{\phi}_p(x)$ denote the convex minorant of $x \mapsto \phi_p(x^{1/3\alpha})$. The assumption that $\frac{\beta}{6}$ is a subdifferential of $\hat{\phi}_p(x)$ at $x = t^{3\alpha}$ implies that the maximum of $\sup_x [\frac{\beta}{6}x - \hat{\phi}_p(x)]$ is attained at $t^{3\alpha}$. Since, for sufficiently small α , the point $(t^{3\alpha}, \phi_p(t))$ lies on $\hat{\phi}_p(x)$, we have that $\hat{\phi}_p(t^{3\alpha}) = \phi_p(t)$ and so the maximum of $\sup_x [\frac{\beta}{6}x - \phi_p(x^{1/3\alpha})]$ is also attained at $t^{3\alpha}$. It follows that the maximum of $\sup_u [\frac{\beta}{6}u^{3\alpha} - \phi_p(u)]$ is attained at t . (However, this maximum may not be unique. If the subtangent line defined by the subdifferential $\frac{\beta}{6}$ touches $\hat{\phi}_p$ at another point $r^{3\alpha}$, then r also a maximum.)

To prove the last part of the lemma, if $\phi_p(u)$ is differentiable at t , then the subdifferential is simply the derivative. Then we have

$$0 = \frac{\partial}{\partial x} \Big|_{x=t^{3\alpha}} [\frac{\beta}{6}x - \phi_p(x^{1/3\alpha})] = \frac{\beta}{6} - \phi'_p(t) \frac{t^{1-3\alpha}}{3\alpha}$$

implies that $\beta = \frac{2\phi'_p(t)}{\alpha t^{3\alpha-1}}$. ■

Next, we use the minorant condition and Lemma 2.9 to define a parameterized family of subregimes of the (p, t) -phase space.

Definition 2.10. Fix $\alpha > 0$. We define the regime \mathcal{S}_α to be the set of pairs (p, t) for which the minorant condition holds with α and there exists a subdifferential $\frac{\beta}{6}$ of the convex minorant of $x \mapsto \phi_p(x^{1/3\alpha})$ such that the variational form $\sup_{0 \leq u \leq 1} [\frac{\beta}{6}u^{3\alpha} - \phi_p(u)]$ is *uniquely* maximized at t . ■

If $\alpha \in [2/3, 1]$, the exponential random graph is known to be asymptotically indistinguishable from an Erdős-Rényi graph $\mathcal{G}_{n,u}$ for some $u \in [0, 1]$. Recalling Definition 2.2 of the replica symmetric phase, the following statement follows directly from the arguments in [16] and Theorem 4.3 in [8].

Lemma 2.11. $\mathcal{S}_{2/3}$ is exactly the replica symmetric phase.

Lemma A.2 shows that $\mathcal{S}_\alpha \supset \mathcal{S}_{\alpha'}$ for $0 < \alpha < \alpha'$. The sets \mathcal{S}_α for $\alpha = 2/3, 1$ are shown in Figure 2.1. Notice also from Figure 2.1 that there exists some a critical value p_{crit} such that when $p \geq p_{crit}$, (p, t) is replica symmetric for all $t \in [p, 1]$; whereas when $p \leq p_{crit}$, there exists an interval $[r_p, \bar{r}_p] \subset (p, 1)$ where (p, t) is replica breaking if $t \in [r_p, \bar{r}_p]$, and (p, t) is replica symmetric for all other values of t .

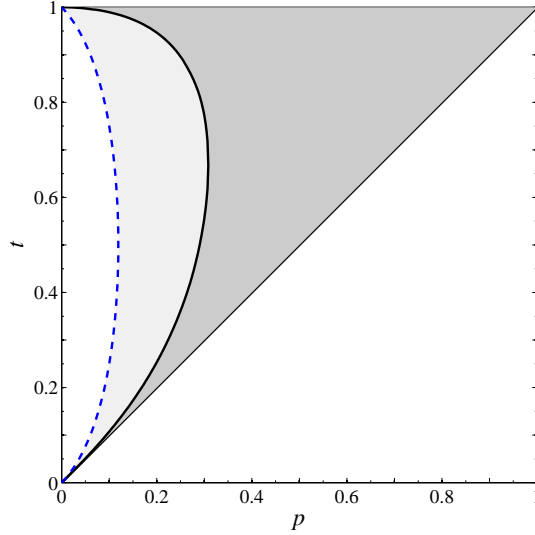


FIGURE 2.1. \mathcal{S}_1 is the dark gray region to the right of the solid curve (not including the solid curve). $\mathcal{S}_{2/3}$, the replica symmetric phase, is the light gray region to the right of the dotted curve (not including the dotted curve), together with the dark gray region. The line is $t = p$.

By definition, any replica symmetric (p, t) satisfies the minorant condition for any $\alpha \in [2/3, 1]$. Are there any replica breaking (p, t) that satisfies the minorant condition for some α ? The answer is in the affirmative. To see this, consider $\alpha = 1/3$ and convex minorant of $x \mapsto \phi_p(x^{1/3\alpha}) = \phi_p(x)$. For each $p < p_{crit}$, there exists an interval $[\underline{r}_p, \bar{r}_p] \subset (p, 1)$ where (p, t) is replica breaking if $t \in [\underline{r}_p, \bar{r}_p]$, and (p, t) is replica symmetric for the other values of t . Since $\phi_p(t) < I_p(t)$ if $t \in [\underline{r}_p, \bar{r}_p]$ and $\phi_p(t) = I_p(t)$ for other values of t , and since $I_p(u)$ is convex, the convex minorant of $\phi_p(x)$ must touch ϕ_p at at least one $t_p \in [\underline{r}_p, \bar{r}_p]$. So (p, t_p) is replica breaking and satisfies the minorant condition.

The preceding argument shows that $\bigcup_{\alpha > 0} \mathcal{S}_\alpha$ is strictly larger than the replica symmetric phase, and contains a nontrivial subset of the replica breaking phase. Using the characterizations of the sets \mathcal{S}_α , we are now ready to define the triangle tilts.

Definition 2.12. Given $(p, t) \in \mathcal{S}_\alpha$ for some $\alpha > 0$, a *triangle tilt with parameter α* corresponding to (p, t) refers to any Gibbs measure $\mathbb{Q}_n^{h_p, \beta, \alpha}$ where $h_p = \log \frac{p}{1-p}$, and $\frac{\beta}{6}$ is any subdifferential of the convex minorant of $x \mapsto \phi_p(x^{1/3\alpha})$. If $\phi_p(u)$ is differentiable at t , then there is exactly one triangle tilt with parameter α corresponding to (p, t) , with the parameters (h_p, β^*, α) where β^* is defined in (2.30). ■

The triangle tilt with parameter α corresponding to (p, t) is well-defined only when $(p, t) \in \mathcal{S}_\alpha$, or, equivalently stated, it is well-defined only when 2.29 holds. In view of Theorem 2.5 and Lemma 2.9, when the triangle tilt with parameter α corresponding to (p, t) is well-defined, it is the measure induced by an exponential random graph which satisfies (2.27) and which is asymptotically indistinguishable from the conditioned Erdős-Rényi graph $\mathcal{G}_{n,p}$ conditioned on the rare event $\{\mathcal{T}(f) \geq t^3\}$. Also, if $(p, t) \in \mathcal{S}_\alpha$, then since by Lemma A.2 the sets $\mathcal{S}_{\alpha'}$ are increasing as α' decreases, the triangle tilt with parameter α' corresponding to (p, t) is defined for any $\alpha' \leq \alpha$. If (p, t) is in the replica symmetric

phase, the triangle tilt can be defined for some $\alpha \in [2/3, 1]$, and since $\phi_p(t) = I_p(t)$ in the replica symmetric phase, from (2.30) the triangle tilt parameters necessarily take on the following explicit expression: (h_p, β^*, α) , where

$$\beta^* = \frac{h_t - h_p}{\alpha t^{3\alpha-1}}. \quad (2.31)$$

If (p, t) is in the replica breaking phase, we may need to resort to numerical strategies to find the parameters β and α .

Remark 2.13. Given any (p, t) , if $\phi_p(u)$ is differentiable at t , then we can define β^* in (2.30) regardless of whether (p, t) belongs to \mathcal{S}_α . In this case, t is a stationary point of the function $L(u) : u \mapsto \frac{\beta}{6}u^{3\alpha} - \phi_p(u)$. If $\phi_p(u)$ is twice differentiable at t , then since

$$\partial_t \beta^* = -\frac{2}{\alpha t^{3\alpha-1}} \left(\frac{\beta\alpha(3\alpha-1)}{2} t^{3\alpha-2} - \phi_p''(t) \right) = -\frac{2}{\alpha t^{3\alpha-1}} L''(t),$$

we have that t is a local maximum of $L(u)$ if and only if $\partial_t \beta^* > 0$.

Now note that in the replica symmetric phase where the LDP implies that an Erdős-Rényi random graph $\mathcal{G}_{n,p}$ conditioned on the event $\{\mathcal{T}(f) \geq t^3\}$ is indistinguishable from $\mathcal{G}_{n,t}$, we have the obvious edge tilt as follows.

Definition 2.14. Given (p, t) , let $h_t = \log \frac{t}{1-t}$. The *edge tilt* refers to the Gibbs measure $\mathbb{Q}_n^{h_t, 0, \alpha} = \mathbb{Q}_n^{h_t, 0} = \mathbb{P}_{n,t}$, corresponding to the Erdős-Rényi graph $\mathcal{G}_{n,t}$.

It is also possible to consider tilts that are a hybrid between the edge tilt and triangle tilts and that satisfy (2.27). Such tilts can be constructed explicitly for the replica symmetric phase. Consider the extremal points of the function

$$V(u) = V(u; h, \beta, \alpha) = \frac{h}{2}u + \frac{\beta}{6}u^{3\alpha} - I(u). \quad (2.32)$$

If the maximum of $V(u)$ occurs at u^* , we must have

$$V'(u^*) = \frac{h}{2} + \frac{\beta\alpha}{2}(u^*)^{3\alpha-1} - \frac{1}{2} \log \frac{u^*}{1-u^*} = 0, \quad (2.33)$$

and

$$V''(u^*) = \frac{\beta\alpha(3\alpha-1)}{2}(u^*)^{3\alpha-2} - \frac{1}{2} \left(\frac{1}{u^*} + \frac{1}{1-u^*} \right) \leq 0. \quad (2.34)$$

Using (2.33) we may express h as a function of β and α :

$$h(\beta, \alpha) = \log \frac{u^*}{1-u^*} - \beta\alpha(u^*)^{3\alpha-1}. \quad (2.35)$$

The next lemma follows from the continuity of V and the conditions (2.33), (2.34).

Lemma 2.15. *Let $u^* \in (0, 1)$ and fix $\alpha \in [2/3, 1]$. For $\beta \geq 0$, let $h(\beta, \alpha)$ be defined by (2.35). There exists $\beta_0 > 0$, depending on α , such that for all $\beta \in [0, \beta_0]$, $V(u; h(\beta, \alpha), \beta, \alpha)$ attains a global maximum uniquely at the point $u = u^*$. In particular, the family of exponential random graphs $\mathcal{G}_n^{h(\beta, \alpha), \beta, \alpha}$ with $\beta \in [0, \beta_0]$ are asymptotically indistinguishable from the Erdős-Rényi graph \mathcal{G}_{n, u^*} with edge probability u^* .*

When (p, t) belongs to the replica symmetric phase, we can apply Lemma 2.15 with $u^* = t$ to obtain a family of hybrid tilts with the parameters $(h(\beta, \alpha), \beta, \alpha)$ for $\beta \in [0, \beta_0]$. Due to Theorem 2.4(b), the hybrid tilt satisfies (2.27) and is asymptotically indistinguishable from the Erdős-Rényi graph $\mathcal{G}_{n,t}$. Hybrid tilts of this form are considered in the numerical simulations in Section 4.2.

3. ASYMPTOTIC OPTIMALITY IN THE REPLICA SYMMETRIC PHASE

The reason for the names, triangle tilt or edge tilt, is that the Radon-Nikodym derivative, $\frac{d\mathbb{P}}{d\mathbb{Q}}$, that weights the samples in the importance sampling estimator (2.3) depends only on the number of triangles or the number of edges, respectively, in the samples. That is,

$$\frac{d\mathbb{P}_{n,p}}{d\mathbb{Q}_{n,h_p,\beta^*,\alpha}}(X) \propto e^{n^2 \frac{\beta^*}{6} \mathcal{T}(X)^\alpha}, \quad \text{and} \quad \frac{d\mathbb{P}_{n,p}}{d\mathbb{Q}_{n,h_t,0}}(X) \propto e^{n^2 \frac{h_t - h_p}{2} \mathcal{E}(X)}.$$

Here recall that $\mathcal{T}(X) = \frac{6}{n^3} T(X)$ is the density of triangles in X and $\mathcal{E}(X) = \frac{2}{n^2} E(X)$ is the density of edges.

In the case of the edge tilt, the fact that the weights depend only on the number of edges has deeper repercussions. Since $\mathbb{E}[\mathcal{E}(\mathcal{G}_n^{h_t,0})] \sim t$, good samples in the target event $\{\mathcal{T}(f) \geq t^3\}$ having fewer than t density of edges are being over-penalized by the weights. In contrast, the triangle tilt penalizes samples more heavily only when they deviate from t^3 density of triangles.

To formalize the above heuristic arguments, we study the asymptotic optimality, or non-optimality, of importance sampling schemes based on the tilted measures $\mathbb{Q}_n^{h,\beta,\alpha}$. For any admissible parameters (h, β, α) , importance sampling estimator based on the tilted measure $\mathbb{Q}_n^{h,\beta,\alpha}$ is

$$\begin{aligned} \tilde{M}_K &= \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{\mathcal{W}_t}(\tilde{X}_k) \frac{d\mathbb{P}_{n,p}}{d\mathbb{Q}_n^{h,\beta,\alpha}}(\tilde{X}_k) \\ &= \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{\mathcal{W}_t}(\tilde{X}_k) \exp \left\{ n^2 \left(\frac{h_p - h}{2} \mathcal{E}(\tilde{X}_k) - \frac{\beta}{6} \mathcal{T}(\tilde{X}_k)^\alpha + \psi_n^{h,\beta,\alpha} - \psi_n^{h_p,0} \right) \right\} \end{aligned} \quad (3.1)$$

where \tilde{X}_k are i.i.d. samples drawn from $\mathbb{Q}_n^{h,\beta,\alpha}$. Denote

$$\hat{q}_n = \hat{q}_n(\tilde{X}) = \mathbf{1}_{\mathcal{W}_t}(\tilde{X}) \frac{d\mathbb{P}_{n,p}}{d\mathbb{Q}_n^{h,\beta,\alpha}}(\tilde{X}).$$

For any (h, β, α) , $\mathbb{E}[\hat{q}_n] = \mu_n$ and so \tilde{M}_K is an unbiased estimator for μ_n .

We now prove the asymptotic optimality of the triangle tilts, Theorem 1.1.

Proof of Theorem 1.1.

Proof. Due to (2.13), it suffices to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}}[\hat{q}_n^2] \leq -2 \inf_{f \in \mathcal{W}_t} \mathcal{I}_p(f). \quad (3.2)$$

Note that $\mathcal{E}, \mathcal{T} : \mathcal{W} \mapsto \mathbb{R}$ are bounded continuous mappings [3, Theorem 3.8], and the exponent of the indicator $\mathbf{1}_{\mathcal{W}_t}(X) = e^{-n^2 \mathbf{0}_{\mathcal{W}_t}(X)}$, where $\mathbf{0}_{\mathcal{W}_t}(X) = 0$ if $X \in \mathcal{W}_t$ and $\mathbf{0}_{\mathcal{W}_t}(X) = \infty$ otherwise, can be approximated by bounded continuous approximations. Since $\mathcal{I}_p(f)$ is the rate function for the family of measures $\mathbb{P}_{n,p}$, (Theorem 3.1 of [5]), we

may apply the Laplace principle: for any (h, β, α) ,

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}_n}[\hat{q}_n^2] &= \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{P}_{n,p}} \left[\mathbf{1}_{\mathcal{W}_t} \frac{d\mathbb{P}_{n,p}}{d\mathbb{Q}_n^{h,\beta,\alpha}} \right] \\
&= - \inf_{f \in \mathcal{W}_t} \left[\mathcal{I}_p(f) + \frac{h - h_p}{2} \mathcal{E}(f) + \frac{\beta}{6} \mathcal{T}(f)^\alpha \right] + \lim_{n \rightarrow \infty} \left(\psi_n^{h,\beta,\alpha} - \psi_n^{h_p,0} \right) \\
&= - \inf_{f \in \mathcal{W}_t} \left[\mathcal{I}_p(f) + \frac{h - h_p}{2} \mathcal{E}(f) + \frac{\beta}{6} \mathcal{T}(f)^\alpha \right] + \tilde{V}(u^*) + \frac{1}{2} \log(1 - p)
\end{aligned} \tag{3.3}$$

where, by (2.23),

$$\tilde{V}(u) := \frac{\beta}{6} u^{3\alpha} - \inf_{f \in \partial \mathcal{W}_u} [\mathcal{I}_p(f)] - \frac{1}{2} \log(1 - p),$$

and $u^* = \operatorname{argsup}_{0 \leq u \leq 1} [\tilde{V}(u)]$. Then

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}}[\hat{q}_n^2] & \\
&= - \inf_{f \in \mathcal{W}_t} \left[\mathcal{I}_p(f) + \frac{h - h_p}{2} \mathcal{E}(f) + \frac{\beta}{6} \mathcal{T}(f)^\alpha \right] + \frac{\beta}{6} (u^*)^{3\alpha} - \inf_{f \in \partial \mathcal{W}_{u^*}} [\mathcal{I}_p(f)] \\
&\leq - \inf_{f \in \mathcal{W}_t} \left[\mathcal{I}_p(f) + \frac{h - h_p}{2} \mathcal{E}(f) \right] + \frac{\beta}{6} ((u^*)^{3\alpha} - t^{3\alpha}) - \inf_{f \in \partial \mathcal{W}_{u^*}} [\mathcal{I}_p(f)]
\end{aligned} \tag{3.4}$$

for any (h, β, α) . The last inequality follows from the fact that $\mathcal{T}(f) \geq t^3$ for all $f \in \mathcal{W}_t$.

Now, taking the triangle tilt with (h_p, β, α) , we have by its definition that $u^* = t$. Then

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}}[\hat{q}_n^2] &\leq - \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)] - \inf_{f \in \partial \mathcal{W}_{u^*}} [\mathcal{I}_p(f)] \\
&= -2 \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]
\end{aligned}$$

Combined with the upper bound for the asymptotic second moment, we conclude that the triangle tilt $\mathbb{Q}_n^{h_p, \beta, \alpha}$ yields an asymptotically optimal importance sampling estimator. ■

3.1. Non-optimality. In this section, we show the non-optimality of importance sampling estimator with certain tilted measures. In the first result, we show that an exponential random graph that is not indistinguishable from the conditioned Erdős-Rényi graph cannot produce an optimal estimator. In the case where (p, t) belongs to the replica symmetric phase, this rules out all exponential random graphs that are indistinguishable from $\mathcal{G}_{n,u}$, with $u \neq t$, from being asymptotically optimal, but does not rule out the Erdős-Rényi graph $\mathcal{G}_{n,t}$ corresponding to the edge tilt. Then, the second non-optimality result identifies a non-trivial subset of the replica symmetric phase for which the edge tilt does not produce an optimal estimator.

Proposition 3.1. *Given (p, t) . Let $\mathcal{G}_n^{h_q, \beta, \alpha}$ be such that the variational form $\sup_{0 \leq u \leq 1} \left[\frac{\beta}{6} u^{3\alpha} - \inf_{f \in \partial \mathcal{W}_u} [\mathcal{I}_q(f)] \right]$ attains its maximum at $u^* \neq t$. Then the importance sampling scheme based on the Gibbs measure tilt $\mathbb{Q}_n^{h_q, \beta, \alpha}$ is not asymptotically optimal.*

Proof. Let f^* be any minimizer of the LDP rate function, $\inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]$. Theorem 2.5 implies that f^* does not maximize $\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)]$. From (3.3) and the first equality

of (2.23),

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}_n} [\hat{q}_n^2] \\
&= - \inf_{f \in \mathcal{W}_t} \left[\mathcal{I}_p(f) + \frac{h - h_p}{2} \mathcal{E}(f) + \frac{\beta}{6} \mathcal{T}(f)^\alpha \right] + \lim_{n \rightarrow \infty} \psi_n^{h, \beta, \alpha} + \frac{1}{2} \log(1 - p) \\
&= - \inf_{f \in \mathcal{W}_t} \left[\mathcal{I}_p(f) + \frac{h - h_p}{2} \mathcal{E}(f) + \frac{\beta}{6} \mathcal{T}(f)^\alpha \right] + \sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] + \frac{1}{2} \log(1 - p) \\
&> - \left[\mathcal{I}_p(f^*) + \frac{h - h_p}{2} \mathcal{E}(f^*) + \frac{\beta}{6} \mathcal{T}(f^*)^\alpha \right] + \frac{h}{2} \mathcal{E}(f^*) + \frac{\beta}{6} \mathcal{T}(f^*)^\alpha - \mathcal{I}(f^*) + \frac{1}{2} \log(1 - p) \\
&= -2\mathcal{T}(f^*) = -2 \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]
\end{aligned}$$

The importance sampling estimator is not asymptotically optimal. \blacksquare

Proposition 3.2. *Let $0 < p < \frac{e^{-1/2}}{1+e^{-1/2}}$ and $t \in (p, 1)$. If t is sufficiently close to 1 and (p, t) belong to the replica symmetric phase, then the importance sampling scheme based on the edge tilt $\mathbb{Q}_n^{h_t, 0}$ is not asymptotically optimal.*

Proof: Starting from (3.4), we have

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}} [\hat{q}_n^2] \\
&= - \inf_{f \in \mathcal{W}_t} \left[\mathcal{I}_p(f) + \frac{\beta}{6} \mathcal{T}(f) + \left(\frac{h(\beta) - h_p}{2} \right) \mathcal{E}(f) \right] - I_p(t) + \frac{\beta}{6} t^3 + \left(\frac{h(\beta) - h_p}{2} \right) t
\end{aligned} \tag{3.5}$$

where $h(\beta) = h(\beta, 1)$ as in (2.35) with $\alpha = 1$. Because (p, t) is in the replica symmetric phase, $\mathcal{I}_p(f)$ is minimized by the constant function

$$f_t(x, y) \equiv t = \arg \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)].$$

On the other hand, \mathcal{E} is minimized by

$$f_1(x, y) = \mathbf{1}_{[0, t]^2}(x, y) = \arg \inf_{f \in \mathcal{W}_t} [\mathcal{E}(f)]. \tag{3.6}$$

This f_1 represents a graph with a large clique, in which there is a complete subgraph on a fraction t of the vertices. Let us define

$$\begin{aligned}
\Gamma(t) &= \mathcal{I}_p(f_t) + \frac{\beta}{6} \mathcal{T}(f_t) + \left(\frac{h(\beta) - h_p}{2} \right) \mathcal{E}(f_t) \\
&= I_p(t) + \frac{\beta}{6} t^3 + \left(\frac{h(\beta) - h_p}{2} \right) t,
\end{aligned}$$

and

$$\begin{aligned}
\Gamma(1) &= \mathcal{I}_p(f_1) + \frac{\beta}{6} \mathcal{T}(f_1) + \left(\frac{h(\beta) - h_p}{2} \right) \mathcal{E}(f_1) \\
&= t^2 I_p(1) + (1 - t^2) I_p(0) + \frac{\beta}{6} t^3 + \left(\frac{h(\beta) - h_p}{2} \right) t^2.
\end{aligned}$$

(Recall $h(\beta) = h_t$ here.) From (3.5) we see that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}} [\hat{q}_n^2] \geq -\Gamma(1) - I_p(t) + \frac{\beta}{6} t^3 + \left(\frac{h(\beta) - h_p}{2} \right) t$$

We claim that for $p < e^{-1/2}/(1 + e^{-1/2})$ and t sufficiently close to 1, we have $\Gamma(1) < \Gamma(t)$. Indeed, let $g(t) = \Gamma(1) - \Gamma(t)$:

$$\begin{aligned} g(t) = \Gamma(1) - \Gamma(t) &= t^2 I_p(1) + (1 - t^2) I_p(0) - I_p(t) + \left(\frac{h_t - h_p}{2} \right) (t^2 - t) \\ &= t^2 I_p(1) + (1 - t^2) I_p(0) - \frac{1}{2} \left(t \log \frac{t}{p} - (1 - t) \log(1 - p) \right) \\ &\quad - \frac{1}{2} (1 - t)^2 \log(1 - t) + \frac{1}{2} \left(\log t - \log \frac{p}{1 - p} \right) (t^2 - t). \end{aligned}$$

Observe that $g(1) = 0$ and

$$g'(1) = 2I_p(1) - 2I_p(0) - 1/2 = -\log \left(\frac{p}{1 - p} \right) - 1/2.$$

So, if $p < e^{-1/2}/(1 + e^{-1/2})$, we have $g'(1) > 0$. So, for t sufficiently close to 1, we have $\Gamma(1) < \Gamma(t)$. Therefore,

$$\Gamma(1) < \Gamma(t) = I_p(t) + \frac{\beta}{6} t^3 + \left(\frac{h(\beta) - h_p}{2} \right) t,$$

and we conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}}[\hat{q}_n^2] &\geq -\Gamma(1) - I_p(t) + \frac{\beta}{6} t^3 + \left(\frac{h(\beta) - h_p}{2} \right) t \\ &> -2I_p(t) = -2 \inf_{f \in \mathcal{W}_t} \mathcal{I}_p(f). \end{aligned}$$

Since the strict inequality holds, the importance sampling scheme associated with $\mathbb{Q}_n^{h,\beta}$ cannot be asymptotically optimal. \blacksquare

Remark 3.3. The critical point in the proposition, $\tilde{p} = \frac{e^{-1/2}}{1 + e^{-1/2}} \approx 0.3775$, corresponds to $h_{\tilde{p}} = -1/2$. In consideration of (2.11) and Figure 2.1, we see that the conditions of the proposition are attainable: if $p < \tilde{p}$ and t is sufficiently close to 1, then (p, t) will be in the replica symmetric phase. For example, when $p = 0.35$, we can numerically approximate the value of $\tilde{t} \approx 0.948$, so that whenever $t \in (\tilde{t}, 1]$, the edge tilt for $(0.35, t)$ is not asymptotically optimal.

4. NUMERICAL SIMULATIONS USING IMPORTANCE SAMPLING

We implement the importance sampling schemes to show the optimality properties of the Gibbs measure tilts in practice. Although we have thus far been considering importance sampling schemes that draw i.i.d. samples from the tilted measure \mathbb{Q} , in practice it is very difficult to sample independent copies of exponential random graphs. This is because of the dependencies of the edges in the exponential random graph, unlike the situation with an Erdős-Rényi graph where the edges are independent. Thus, to implement the importance sampling scheme, we turn to a Markov chain Monte Carlo method known as the Glauber dynamics to generate samples from the exponential random graph. The Glauber dynamics refers to a Markov chain whose stationary distribution is the Gibbs measure $\mathbb{Q}_n^{h,\beta,\alpha}$. The samples \tilde{X}_k from the Glauber dynamics are used to form the importance sampling estimator \tilde{M}_K in (3.1). The variance of \tilde{M}_K clearly also depends on the correlation between the successive samples. However, in this paper, rather than focus on the effect of correlation on the variance of \tilde{M}_K , we instead investigate and compare the

optimality of the importance sampling schemes, and show that importance sampling is a viable method for moderate values of n .

4.1. Glauber dynamics. For the exponential random graph $\mathcal{G}_n^{h,\beta,\alpha}$, the Glauber dynamics proceeds as follows.

Suppose we have a graph $X = (X_{ij})_{1 \leq i < j \leq n}$. The graph \tilde{X} is generated from X via the following procedure.

1. Choose an edge X_{ij} , for some (i, j) , from X uniformly at random.
2. For the new graph \tilde{X} , fix all other edges $\tilde{X}_{i'j'} = X_{i'j'}$, for $(i', j') \neq (i, j)$.
3. Conditioned on all other edges fixed, pick

$$\tilde{X}_{ij} \sim \text{Bern}(\varphi)$$

where

$$\varphi = \frac{e^{h+(\beta/n)(L_{ij}+M_{ij})^\alpha(n^3/6)^{1-\alpha}}}{e^{h+(\beta/n)(L_{ij}+M_{ij})^\alpha(n^3/6)^{1-\alpha}} + e^{(\beta/n)M_{ij}^\alpha(n^3/6)^{1-\alpha}}}$$

and where

$$L_{ij} = \sum_{k \neq i,j} X_{ik}X_{jk}, \quad \text{and} \quad M_{ij} = \sum_{(k,l,m) \not\supseteq (i,j)} X_{kl}X_{km}X_{lm},$$

is the number of 2-stars in X with a base at the edge X_{ij} , and the number of triangles in X not involving the edge X_{ij} , respectively.

4. If conditioning on A_J is used, check if \tilde{X} is in A_J . If not, revert to X .

In step 4, a conditioning of the Gibbs measure is discussed in Section 4.3.

For the classical exponential random graph with $\alpha = 1$, the probability φ in the Glauber dynamics has a neater expression,

$$\varphi = \frac{e^{h+\beta L_{ij}/n}}{1 + e^{h+\beta L_{ij}/n}}.$$

At each MCMC step, if $X_{ij} \neq \tilde{X}_{ij}$, then $E(\tilde{X})$ differs from $E(X)$ by one edge, and $T(\tilde{X})$ differs from $T(X)$ by nL_{ij} triangles. The stationary distribution of the Glauber dynamics is the Gibbs measure $\mathbb{Q}_n^{h,\beta,\alpha}$ that defines the exponential random graph $\mathcal{G}_n^{h,\beta,\alpha}$. Regarding the mixing time of the Markov chain, [1] showed that if $(h, \beta, 1)$ has the property that the unique global maximum of the function $V(u)$, defined in (2.32), is also the unique turning point, then the mixing time for the Glauber dynamics is $\mathcal{O}(n^2 \log n)$. For other values of $(h, \beta, 1)$, the mixing time is $\mathcal{O}(e^n)$.

4.2. Numerical simulations in the replica symmetry phase. The importance sampling scheme was performed for $p = 0.35, t = 0.4$, in the replica symmetry phase, using the Glauber dynamics. The simulations used the Gibbs measure tilts $\mathbb{Q}_n^{h(\beta),\beta}$, with $\alpha = 1$ and β is of the form

$$\beta = \beta_q = \frac{1}{t^2} \left(\log \frac{t}{1-t} - \log \frac{q}{1-q} \right), \quad \text{for } q = 0.35, 0.36, \dots, 0.4 \quad (4.1)$$

and $h(\beta_q)$ is given by (2.35). Each of these exponential random graphs $\mathcal{G}_n^{h(\beta),\beta_q}$ is indistinguishable from the Erdős-Rényi graph $\mathcal{G}_{n,t}$, and $q = p = 0.35$ is the triangle tilt while $q = t = 0.4$ is the edge tilt. For the values $p \leq q \leq t$, the Table 4.1 verifies the accuracy of the importance sampling estimates for $\mu_n := \mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{W}_t)$ using the tilts $\mathbb{Q}_n^{h(\beta_q),\beta_q}$. Also

$n \setminus q$	0.35	0.36	0.37	0.38	0.39	0.4
16	0.12475 (-0.008131)	0.1247 (-0.008132)	0.12521 (-0.008116)	0.12425 (-0.008146)	0.12441 (-0.008141)	0.12435 (-0.008143)
32	0.01107 (-0.004398)	0.011056 (-0.004399)	0.011116 (-0.004394)	0.010941 (-0.004409)	0.010972 (-0.004407)	0.010729 (-0.004429)
64	2.1919e-06 (-0.003181)	2.0283e-06 (-0.003200)	2.6073e-06 (-0.003139)	5.3287e-07 (-0.003527)	1.3822e-06 (-0.003294)	3.5772e-06 (-0.003062)
96	1.1036e-11 (-0.002738)	1.6868e-11 (-0.002692)	2.0805e-11 (-0.002669)	4.4039e-11 (-0.002587)	2.6124e-11 (-0.002644)	4.497e-11 (-0.002585)

TABLE 4.1. Comparison of the estimates for the probability μ_n (top number) for varying tilts $\mathbb{Q}_n^{h(\beta),\beta}$, where the parameters $\beta = \beta_{q,t}$ are defined in (4.1). Also shown is the log probability $\frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{W}_t)$ (lower number).

$n \setminus q$	0.35	0.36	0.37	0.38	0.39	0.4
16	0.030839 (-0.01199)	0.03007 (-0.01206)	0.030173 (-0.01204)	0.030553 (-0.01203)	0.031902 (-0.01191)	0.034105 (-0.01173)
32	0.00039386 (-0.007391)	0.00038614 (-0.007407)	0.00040055 (-0.007377)	0.00042058 (-0.007347)	0.00047462 (-0.007253)	0.00052598 (-0.00718)
64	2.9982e-11 (-0.005879)	2.5716e-11 (-0.005917)	4.6804e-11 (-0.005774)	2.3144e-12 (-0.006513)	1.9783e-11 (-0.005995)	1.8035e-10 (-0.005461)
96	1.157e-21 (-0.005220)	2.7721e-21 (-0.005125)	4.9044e-21 (-0.005065)	2.8661e-20 (-0.004876)	1.4562e-20 (-0.004951)	6.8628e-20 (-0.004785)

TABLE 4.2. Comparison of the estimates for the variance $Var_{\mathbb{Q}}(\hat{q}_n)$ (top number) for varying tilts $\mathbb{Q}_n^{h(\beta),\beta}$, where the parameters $\beta = \beta_{q,t}$ are defined in (4.1). Also shown is the log second moment $\frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}}[\hat{q}_n^2]$ (lower number).

shown is the estimate for the log probability, $\frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{W}_t)$. Since (p, t) is replica symmetric, the LDP rate is

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{W}_t) = -I_p(t) = -0.002694.$$

The value of the log probability is seen to approach the LDP rate as n is increased.

Table 4.2 shows the estimated values of the variance of the estimator, $Var_{\mathbb{Q}_n}(\hat{q}_n)$, where $\hat{q}_n = \mathbf{1}_{\mathcal{W}_t} \frac{d\mathbb{P}_{n,p}}{d\mathbb{Q}_n^{h,\beta}}$, as well as the log second moment $\frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}_n}[\hat{q}_n^2]$. The variance of the estimator for all the tilts appear to be comparable and the log second moment likewise appears to converge towards $-2I_p(t) = -0.0053869$. In this case, $p = 0.35, t = 0.4$ does not belong to the regime described in Proposition 3.2, and the numerical results suggest that all the tilts of this form, including the edge tilt, appear to be close to asymptotically optimal.

For $n = 16, 32, 64$, the number of MCMC samples used was $5 \times 10^4 n^2 \log n$, while for $n = 96$, the number of MCMC samples used was $10^5 n^2 \log n$.

Both the random graphs corresponding to the triangle or edge tilts are expected by (2.27), (2.28) to have $\binom{n}{3} t^3$ triangles and $\binom{n}{2} t$ edges on average. However, there is a difference between the way that the triangle and edge tilts produce events in $\{\mathcal{T}(f) \geq t^3\}$, and that is in the number of edges in the successful samples that fall in the rare event.

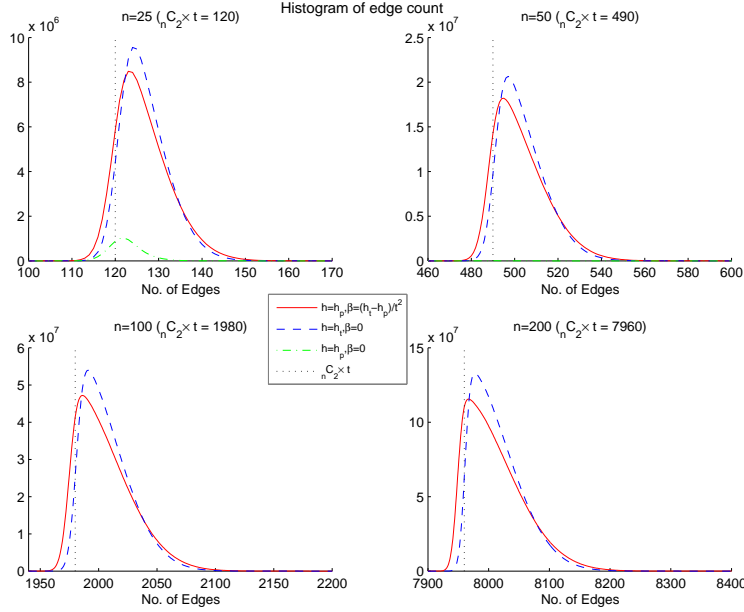


FIGURE 4.1. Histogram of edge counts in the samples obtained by the triangle and edge tilts, conditioned on the rare event. The solid red line shows the triangle tilt; the dashed blue line shows the edge tilt; the dotted green line shows the Monte Carlo sampling.

The distribution of the edge count of successful samples from the triangle tilt has a larger proportion with less than $\binom{n}{2}t$ edges, as compared to the edge tilt. This is shown in Figure 4.1.

4.3. Importance sampling with conditioned Gibbs measures. Quite a different issue from the asymptotic optimality of the importance sampling estimator is the question of the efficiency of the Glauber dynamics in drawing samples from the tilted measure. The efficiency of using an MCMC to draw samples is subject to the mixing time of the Markov chain. In the case of the exponential random graph, the mixing time of some such graphs is known to be exponentially long, $\mathcal{O}(e^n)$, due to the fact that the Hamiltonian $\mathcal{H}(f)$ has multiple local maxima [1]. In this section, we propose a way to sidestep this issue, by using a conditioned version of the Gibbs measure, in which the sampling from the exponential random graph is restricted to an appropriate subregion of the state space Ω_n . Conditioning the Gibbs measure on the desired subregion of the state space serves to focus the sampling to the region of the state space that really matters, and possibly also improving the mixing time of the Markov chain.

The conditioned Gibbs measure is particularly apt in the following scenario. Suppose, for given (p, t) , the variation form in (2.29) is locally, but not globally, maximized by t (c.f. Figure 4.2). If $u^* \neq t$ is the global maximum of (2.29), then $\mathcal{G}_n^{h, \beta, \alpha}$ is indistinguishable from the conditioned Erdős-Rényi graph $\mathcal{G}_{n, p}$ conditioned on $\{\mathcal{T}(f) \geq (u^*)^3\}$. Compared to our target of exceeding $\binom{n}{3}t^3$ triangles, the samples from $\mathcal{G}_n^{h, \beta, \alpha}$ will have an over- or under-abundance of triangles, leading to a poor estimator with very large variance. Recall that Proposition 3.1 shows that the importance sampling estimator based on $\mathbb{Q}_n^{h, \beta, \alpha}$ is

non-optimal. The conditioned Gibbs measure mitigates this problem by restricting the exponential random graph to having just the “right” number of triangles.

Conditioned Gibbs measure. Given a set $A \subset \mathcal{W}$, the exponential random graph conditioned on A , denoted $\mathcal{G}_{n,A}^{h,\beta,\alpha}$ has the conditional Gibbs measure $\tilde{\mathbb{Q}}_{n,A}^{h,\beta,\alpha} = \mathbb{Q}_n^{h,\beta,\alpha}|_A$,

$$\tilde{\mathbb{Q}}_{n,A}^{h,\beta,\alpha}(X) \propto \begin{cases} e^{n^2 \mathcal{H}(X)}, & \text{if } X \in A \\ 0 & \text{if } X \notin A \end{cases}$$

where the Hamiltonian $\mathcal{H}(X)$ is defined in (2.16). The free energy $\tilde{\psi}_{n,A} = \tilde{\psi}_{n,A}^{h,\beta,\alpha}$ is

$$\tilde{\psi}_{n,A} = \frac{1}{n^2} \log \sum_{X \in A} e^{n^2 \mathcal{H}(X)}$$

The following proposition describes the asymptotic behaviour of the free energy, which is analogue of Theorems 3.1 and 3.2 in [5].

Proposition 4.1. *For any bounded continuous mapping $\mathcal{H} : \mathcal{W} \mapsto \mathbb{R}$, and any closed subset $A \subset \mathcal{W}$, let $\tilde{\psi}_{n,A} = \frac{1}{n^2} \log \sum_{X \in A} e^{n^2 \mathcal{H}(X)}$. Then*

$$\lim_{n \rightarrow \infty} \tilde{\psi}_{n,A} = \sup_{f \in \mathcal{W} \cap A} [\mathcal{H}(f) - \mathcal{I}(f)]. \quad (4.2)$$

Moreover, if $\tilde{\mathcal{F}} \subset \mathcal{W}$ is the subset on which $[\mathcal{H}(f) - \mathcal{I}(f)]$ is maximized, then for any ϵ , there exists a constant $\delta > 0$ such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\delta_{\square}(\mathcal{G}_{n,A}, \tilde{\mathcal{F}}) > \epsilon) \leq \delta.$$

Proof. This follows from a simple modification of the proof of Theorem 3.1 and 3.2 in [5] to restrict to the set A . ■

The importance sampling scheme based on the conditioned Gibbs measure $\tilde{\mathbb{Q}}_{n,A}^{h,\beta,\alpha}$ gives the estimator

$$\hat{\nu}_A = \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{\mathcal{W}_t}(\tilde{X}_k) \frac{d\tilde{\mathbb{P}}_{n,p,A}}{d\tilde{\mathbb{Q}}_{n,A}^{h,\beta,\alpha}}(\tilde{X}_k), \quad \text{where } \tilde{X}_k \sim \text{i.i.d. } \tilde{\mathbb{Q}}_{n,A}^{h,\beta,\alpha} \quad (4.3)$$

where $\tilde{\mathbb{P}}_{n,p,A} = \mathbb{P}_{n,p}|_A$. Note that $\hat{\nu}_A$ is an unbiased estimator for $\nu_{n,p} = \tilde{\mathbb{P}}_{n,p,A}(\mathcal{W}_t)$. The estimator $\hat{\mu}_n$ for $\mu_n = \mathbb{P}_{n,p}(\mathcal{W}_t)$ can be obtained from $\hat{\nu}_A$ by

$$\hat{\mu}_n = \hat{\nu}_A \cdot \mathbb{P}_{n,p}(A) + \mathbb{P}_{n,p}(\mathcal{W}_t \cap A^c).$$

Since the two probabilities on the RHS, particularly the second term, may not be easily computable or estimated, we may alternatively take $\hat{\nu}_A$ as a biased estimator for μ_n . By an appropriate choice of the set A , we can ensure the bias is small and vanishes exponentially faster than the small probability we are trying to estimate (see Lemma A.4(ii)).

In our application, the conditioning of the Gibbs measure is applied to control the number of triangles that the sampled graphs are allowed to have. Thus, it is natural to choose the set A of the form

$$A_J = \{f \in \mathcal{W} : \mathcal{T}(f) \in J\} \cap A_0 \quad (4.4)$$

where $J \subset [0, 1]$ is a closed interval and $A_0 \subset \mathcal{W}$ is a closed subset containing all the constant functions in J . Then the set $A_J \subset \mathcal{W}$ is closed because \mathcal{T} is continuous in the cut distance metric δ_{\square} . A consequence of Proposition 4.1 is a variational formulation similar to (2.23).

Proposition 4.2. *Let A_J be defined in (4.4). Given any Gibbs measure parameters (h, β, α) , assume wlog that $h = h_q = \log \frac{q}{1-q}$ for some $q \in (0, 1)$. For $u \in [0, 1]$, denote $\partial\mathcal{W}_u := \{f \in \mathcal{W} \mid \mathcal{T}(f) = u^3\}$ and let $\mathcal{F}_u^* \subset \mathcal{W}$ be the set of minimizers of $\inf_{f \in A_0 \cap \partial\mathcal{W}_u} [\mathcal{I}_q(f)]$. Then*

$$\lim_{n \rightarrow \infty} \tilde{\psi}_{n, A_J} = \sup_{f \in A_J} [\mathcal{H}(f) - \mathcal{I}(f)] = \sup_{u \in J} \left[\frac{\beta}{6} u^{3\alpha} - \inf_{f \in A_0 \cap \partial\mathcal{W}_u} [\mathcal{I}_q(f)] - \frac{1}{2} \log(1 - q) \right] \quad (4.5)$$

The supremum $\sup_{f \in A_J} [\mathcal{H}(f) - \mathcal{I}(f)]$ is attained exactly on the set $\mathcal{F}_{v^*}^*$, where v^* maximizes the RHS of (4.5).

Further, if (q, v^*) belongs to the replica symmetric phase, then the supremum $\sup_{f \in A_J} [\mathcal{H}(f) - \mathcal{I}(f)]$ is attained uniquely by the constant function $f_{v^*}^* \equiv v^*$, and

$$\lim_{n \rightarrow \infty} \tilde{\psi}_{n, A_J} = V(v^*) = \sup_{u \in J} [V(u)].$$

where $V(u) = \frac{h}{2}u + \frac{\beta}{6}u^{3\alpha} - I(u)$.

The proof is identical to the proof of Theorem 2.5 and is left to the appendix. Combining Propositions 4.1 and Theorem 4.2, the exponential random graph conditioned on A_J is asymptotically indistinguishable from the graphs in the set $\mathcal{F}_{v^*}^*$. The case when $J = [0, 1]$ and $A_0 = \mathcal{W}$, which is when there is no conditioning, coincides with Theorem 2.5.

Using Proposition 4.2, the notion of the triangle tilt can be extended to the importance sampling schemes using the conditioned Gibbs measures, in a similar way as in Section 2.2 for the full Gibbs measure, as follows. Given (p, t) and the set A_J , suppose there exists parameters (h_p, β, α) such that

$$t = \operatorname{argsup}_{u \in J} \left[\frac{\beta}{6} u^{3\alpha} - \inf_{f \in \partial\mathcal{W}_u} [\mathcal{I}_p(f)] \right]. \quad (4.6)$$

The conditioned Gibbs measure $\mathcal{G}_{n, A_J}^{h_p, \beta, \alpha}$ is a (conditioned) triangle tilt with parameter α corresponding to (p, t) .

Under some mild conditions on the set A_J , Lemma A.4 shows that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n, p, A_J}(\mathcal{W}_t) = - \inf_{f \in \mathcal{W}_t \cap A_J} [\mathcal{I}_p(f)] = - \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]. \quad (4.7)$$

Thanks to (4.7), the notion of asymptotic optimality for conditioned tilts is unchanged. As a corollary, we have that the conditioned triangle tilt also yields an asymptotically optimal importance sampling scheme. The proof is left to the appendix.

Corollary 4.3. *Given any (p, t) , let A_J be defined in (4.4) with $p \in J$ and $t \in J^\circ$ in the interior of J . Suppose that there exists a conditioned Gibbs measure $\mathcal{G}_{n, A_J}^{h_p, \beta, \alpha}$ is a conditioned triangle tilt $\mathbb{Q}_n^{h_p, \beta, \alpha}$ satisfying (4.6). Then the importance sampling estimator based on the conditioned triangle tilt is asymptotically optimal.*

We remark here that if the Glauber dynamics is used to generate samples from $\mathbb{Q}_{n, A_J}^{h_p, \beta, \alpha}$, we must require that A_J is connected. A sufficient condition for A_J to be connected is if J is an interval of the form $[0, r]$ or $[r, 1]$.

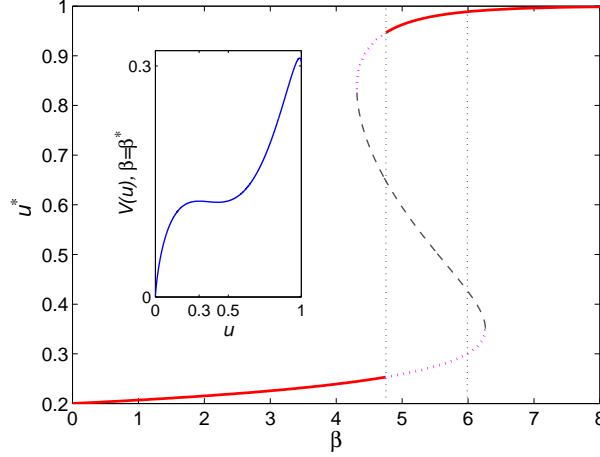


FIGURE 4.2. The phase curve denotes the values of the stationary points of the variational form $V(u) = \frac{h}{2}u + \frac{\beta}{6}u^{3\alpha} - I(u)$, as β varies, and given $\alpha = 1$, $h_p = \log \frac{p}{1-p}$, $p = 0.2$. The red solid line denotes when the stationary point is a global maximum of $V(u)$; the red dotted line denotes the local maximum; the blue dashed line denotes the local minimum. At the phase transition point at $\beta \approx 4.76$, the maximum of the variational form jumps from $u^* \approx 0.253$ to $u^* \approx 0.947$. The inset shows the function $V(u)$ for $\beta = \beta^* \approx 5.99$ attaining a local maximum at $t = 0.3$ and global maximum at $u^* \approx 0.989$.

Numerical illustration of a conditioned Gibbs measure. We illustrate the conditional Gibbs measure tilt with an example. For concreteness, let us set $p = 0.2$ and $t = 0.3$, and for the Gibbs measure parameters, set $h = h_p$ and $\alpha = 1$, and vary $\beta \geq 0$. We will study how the asymptotic second moment changes as β varies. The pair $(p, t) = (0.2, 0.3)$ is in the replica symmetric phase $\mathcal{S}_{2/3}$. For the triangle tilt with $\alpha = 1$, we have from (2.31) $\beta = \beta^* = (h_t - h_p)/t^2$. The variational form $V(u; h_p, \beta^*) = \frac{h_p}{2}u + \frac{\beta^*}{6}u^3 - I(u)$ has a local maximum at $t = 0.3$ but is maximized at a value $u^* \approx 0.989$. (See Figure 4.2.) So $(p, t) \notin \mathcal{S}_1$. The exponential graph $\mathcal{G}_n^{h_p, \beta^*}$ will produce on average $\binom{n}{3}(u^*)^3$ triangles—this is too many triangles, and the variance of the importance sampling estimator will blow up.

To avoid getting samples with too many triangles, let us restrict the state space to cap the number of triangles and edges,

$$A_r = \{f \in \mathcal{W} : \mathcal{T}(f) \leq r^3 \text{ and } \mathcal{E}(f) \leq r\} \quad (4.8)$$

for some $r > t$. With $h = h_p$ fixed and for $\beta > 0$, the asymptotic second moment of the estimator under $\tilde{\mathbb{Q}}_n = \tilde{\mathbb{Q}}_{n, A_r}^{h_p, \beta}$ is

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\tilde{\mathbb{Q}}_n} [\hat{q}_{n, A_r}^2] &= - \inf_{f \in \mathcal{W}_t \cap A_J} \left[\mathcal{I}_p(f) + \frac{\beta}{6} \mathcal{T}(f) \right] + V(u_r^*) + \frac{1}{2} \log(1-p) \\ &= - \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)] + \frac{\beta}{6} ((u_r^*)^3 - t^3) - \inf_{f \in \partial \mathcal{W}_{u_r^*}} [\mathcal{I}_p(f)] \end{aligned}$$

where $u_r^* = \operatorname{argsup}_{0 \leq u \leq r} [V(u; h_p, \beta)]$.

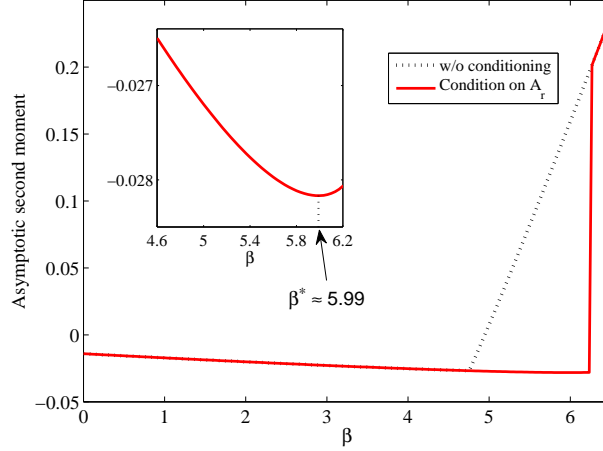


FIGURE 4.3. A plot of the asymptotic second moment, $\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}}[\hat{q}_{n,A_J}^2]$, of the importance sampling estimator based on the conditioned Gibbs tilt for fixed $h = h_p$ and varying β . The insert is a zoom-in to show that the smallest variance is attained at $\beta = \beta^*$. The dotted line shows the rapid deterioration of the asymptotic second moment of the estimator without the use of conditioning. Parameters used are $p = 0.2$ and $t = 0.3$.

Figure 4.3 shows the asymptotic second moment for the tilts (h_p, β) both with and without conditioning of the Gibbs measure.

When the tilt with $\beta = \beta^*$ is conditioned on A_r , it gives the best estimator and is asymptotically optimal by Corollary 4.3. This is corroborated by the numerical simulations that suggest that the triangle tilt performs significantly better than crude Monte Carlo sampling, and also outperforms the edge tilt. In contrast, when no conditioning is performed, the IS estimator exhibits a sharp decline in performance when β is increased beyond the transition point at $\beta \approx 4.76$ (c.f. Figure 4.2). This transition point coincides with the phase transition when the exponential graph $\mathcal{G}_n^{h_p, \beta}$ exhibits a transition from a graph with low edge density to one with high edge density. As mentioned above, the graph with high edge density overproduces triangles, causing the estimator to have a large variance.

Triangle tilt with parameter α and conditioned Gibbs measure. The importance sampling scheme was next performed for $p = 0.2$ and $t = 0.3$, in the replica symmetric phase. We now consider the following tilted measures, all of whose exponential random graphs are indistinguishable from the Erdős-Rényi graph $\mathcal{G}_{n,t}$.

- Triangle tilt with $\alpha = 2/3$: $\mathbb{Q}_n^{h_p, \beta_{2/3}^*, 2/3}$, where $\beta_{2/3}^*$ is defined in (2.31) with $\alpha = 2/3$.
- Conditioned triangle tilt with $\alpha = 1$: $\mathbb{Q}_{n, A_r}^{h_p, \beta_1^*}$, where β_1^* is defined in (2.31) with $\alpha = 1$, and A_r is defined in (4.8) with $r \approx 0.4272 > t$, which is a local minimum of $V(u)$.
- Edge tilt: $\mathbb{Q}_n^{h_t, 0} = \mathbb{P}_{n,t}$.

n	Triangle tilt $\alpha = 2/3$	Conditioned triangle tilt	Edge tilt	Monte Carlo
16	0.0064 (-0.0197)	0.006474 (-0.0197)	0.006285 (-0.0198)	
32	4.3148e-7 (-0.0143)	3.5488e-7 (-0.0145)	3.3878e-7 (-0.0145)	3.7758e-7 (-0.0144)
48	1.3976e-13 (-0.0128)	1.1418e-14 (-0.0139)	1.2039e-12 (-0.0119)	— —
64	6.1882e-21 (-0.0136)	2.9076e-23 (-0.0127)	1.8316e-19 (-0.0105)	— —

TABLE 4.3. Estimates for the probability μ_n . In parenthesis is the estimator for the log probability $\frac{1}{n^2} \log \mu_n$.

n	Triangle tilt $\alpha = 2/3$	Conditioned triangle tilt	Edge tilt	Monte Carlo
16	1.5059e-4 (-0.0334)	2.4166e-4 (-0.0319)	1.0391e-3 (-0.0267)	
32	1.5222e-12 (-0.0265)	1.9083e-12 (-0.0263)	6.7116e-11 (-0.0229)	3.7758e-7 (-0.0144)
48	2.6058e-25 (-0.0245)	3.268e-27 (-0.0265)	2.4737e-20 (-0.0196)	— —
64	7.1703e-40 (-0.0220)	2.8806e-44 (-0.0245)	1.2806e-33 (-0.0185)	— —

TABLE 4.4. Estimates for the variance $\text{Var}_{\mathbb{Q}_n}(\hat{q}_n)$. In parenthesis is the estimate for the log second moment, $\frac{1}{n^2} \log \mathbb{E}_{\mathbb{Q}_n}[\hat{q}_n^2]$.

Tables 4.3, 4.4 shows the estimated values for the mean and variance of $\hat{q}_n = \mathbf{1}_{\mathcal{W}_t} \frac{d\tilde{\mathbb{P}}_{n,p,AJ}}{d\tilde{\mathbb{Q}}_{n,AJ}^{h,\beta}}$. The direct Monte Carlo simulation is shown for $n = 32$ to verify the estimates. We observe that both triangle tilts perform comparably, and both outperform the edge tilt.

APPENDIX A. AUXILIARY LEMMAS AND PROOFS

We collate a number of lemmas and proofs in this section, roughly in the order that they appear in the paper.

Lemma A.1. (i) Given (p, t) , let \mathcal{F}^* be the set of functions that minimize the LDP rate function, $\inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]$ in (2.9). Then \mathcal{F}^* is the minimal set that the Erdős-Rényi graph $\mathcal{G}_{n,p}$ conditioned on $\{\mathcal{T}(f) \geq t^3\}$ is asymptotically indistinguishable from.
(ii) Given (h, β, α) , let \mathcal{F}^* be the set of functions that maximize $\sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)]$. Then \mathcal{F}^* is the minimal set that the exponential random graph $\mathcal{G}_n^{h,\beta,\alpha}$ is asymptotically indistinguishable from.

Proof. The proofs of asymptotic indistinguishability of \mathcal{F}^* was shown in [8, Theorem 3.1] for (i) and [5, Theorem 3.22] for (ii). The proofs naturally extend to give the minimality of \mathcal{F}^* , and we state them here for the record.

Observe that for any random graph \mathcal{G}_n that is asymptotically indistinguishable from a set \mathcal{F}^* , to show that \mathcal{F}^* is minimal, it suffices to show that, for any relatively open

non-empty subset $\mathcal{F}_0 \subset \mathcal{F}^*$ such that $\mathcal{F}^* \setminus \mathcal{F}_0$ is non-empty, there exists $\epsilon > 0$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\delta_{\square}(\mathcal{G}_n, \mathcal{F}^* \setminus \mathcal{F}_0) > \epsilon) = 0. \quad (\text{A.1})$$

Let $\mathcal{F}_0 \subset \mathcal{F}^*$ be any relatively open non-empty subset, with $\mathcal{F}^* \setminus \mathcal{F}_0$ non-empty. Denote, for $\epsilon > 0$,

$$\mathcal{F}_{\epsilon} = \{f \in \mathcal{W} \mid \delta_{\square}(f, \mathcal{F}^* \setminus \mathcal{F}_0) > \epsilon\}.$$

(i) Since \mathcal{F}_0 is relatively open in \mathcal{F}^* , $\delta_{\square}(f, \mathcal{F}^* \setminus \mathcal{F}_0) > 0$ for any $f \in \mathcal{F}_0$. So, there exists an $\epsilon > 0$ sufficiently small such that $(\mathcal{F}_{\epsilon} \cap \mathcal{W}_t)^{\circ}$ contains at least one element of \mathcal{F}_0 . (A° denotes the interior of A .) It follows that

$$\inf_{f \in (\mathcal{F}_{\epsilon} \cap \mathcal{W}_t)^{\circ}} [\mathcal{I}_p(f)] = \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)].$$

Since

$$\mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{F}_{\epsilon} \mid \mathcal{G}_{n,p} \in \mathcal{W}_t) = \frac{\mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{F}_{\epsilon} \cap \mathcal{W}_t)}{\mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{W}_t)},$$

from the large deviation principle in [8, Theorem 2.3] implies that

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{F}_{\epsilon} \mid \mathcal{G}_{n,p} \in \mathcal{W}_t) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{F}_{\epsilon} \cap \mathcal{W}_t) - \frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{W}_t) \\ &\geq - \inf_{f \in (\mathcal{F}_{\epsilon} \cap \mathcal{W}_t)^{\circ}} [\mathcal{I}_p(f)] + \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)] \\ &= 0. \end{aligned}$$

(ii) Since \mathcal{F}_0 is relatively open in \mathcal{F}^* , there exists an $\epsilon > 0$ sufficiently small such that $\mathcal{F}_{\epsilon}^{\circ}$ contains at least one element of \mathcal{F}_0 , and

$$\inf_{f \in \mathcal{F}_{\epsilon}^{\circ}} [\mathcal{H}(f) - \mathcal{I}(f)] = \inf_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)].$$

Since the Hamiltonian \mathcal{H} is bounded, for any $\eta > 0$, there is a finite set $A \subset \mathbb{R}$ such that the intervals $\{(a, a + \eta), a \in A\}$ cover the range of \mathcal{H} . Let $\mathcal{F}_{\epsilon}^a = \mathcal{F}_{\epsilon} \cap \mathcal{H}^{-1}([a, a + \eta])$, and let $\mathcal{F}_{\epsilon}^{a,n} = \mathcal{F}_{\epsilon}^a \cap \Omega_n$ be the functions corresponding to a simple finite graph. Then

$$\mathbb{P}(\mathcal{G}_n \in \mathcal{F}_{\epsilon}) \geq \sum_{a \in A} e^{n^2(a - \psi_n)} |\mathcal{F}_{\epsilon}^{a,n}| \geq e^{-n^2 \psi_n} \sup_{a \in A} [e^{n^2 a} |\mathcal{F}_{\epsilon}^{a,n}|]$$

and

$$\frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_n \in \mathcal{F}_{\epsilon}) \geq -\psi_n + \sup_{a \in A} [a - \frac{1}{n^2} \log |\mathcal{F}_{\epsilon}^{a,n}|].$$

By an observation in [5, Eqn. (3.4)], for any open set $U \subset \mathcal{W}$, and $U_n = U \cap \Omega_n$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log |U_n| \geq - \inf_{f \in U} [\mathcal{I}(f)].$$

Then, since

$$\sup_{f \in \mathcal{F}_{\epsilon}^a} [\mathcal{H}(f) - \mathcal{I}(f)] \leq \sup_{f \in \mathcal{F}_{\epsilon}^a} [a + \eta - \mathcal{I}(f)] = a + \eta - \inf_{f \in \mathcal{F}_{\epsilon}^a} [\mathcal{I}(f)]$$

we have that

$$\begin{aligned}
\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}(\mathcal{G}_n \in \mathcal{F}_\varepsilon) &\geq - \sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] + \sup_{a \in A} [a - \inf_{f \in (F_\varepsilon^a)^\circ} [\mathcal{I}(f)]] \\
&\geq - \sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] + \sup_{a \in A} \sup_{f \in (F_\varepsilon^a)^\circ} [\mathcal{H}(f) - \mathcal{I}(f)] - \eta \\
&\geq - \sup_{f \in \mathcal{W}} [\mathcal{H}(f) - \mathcal{I}(f)] + \sup_{f \in F_\varepsilon^\circ} [\mathcal{H}(f) - \mathcal{I}(f)] - \eta \\
&= 0.
\end{aligned}$$

The proof is complete. ■

Proof of Proposition 2.8.

Proof. Let $\epsilon_1 > 0$ be arbitrary. As in Theorem 2.5, let $\mathcal{F}_{v^*}^*$ be the set of minimizers of $\inf_{f \in \partial \mathcal{W}_{v^*}} [\mathcal{I}_q(f)]$.

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_n} |\mathcal{T}(X) - (v^*)^3| \\
&= \int_{\{\delta_\square(X, \mathcal{F}_{v^*}^*) > \epsilon_1\}} |\mathcal{T}(X) - (v^*)^3| d\mathbb{Q}_n(X) + \int_{\{\delta_\square(X, \mathcal{F}_{v^*}^*) \leq \epsilon_1\}} |\mathcal{T}(X) - (v^*)^3| d\mathbb{Q}_n(X) \\
&= (I) + (II)
\end{aligned}$$

(We have dropped the superscripts, $\mathbb{Q}_n = \mathbb{Q}_n^{h_q, \beta, \alpha}$.) We estimate the two terms.

To estimate (I), by [5, Theorem 4.2], there exists $C, \epsilon_2 > 0$ such that for sufficiently large n

$$\mathbb{Q}_n(\delta_\square(X, \mathcal{F}_{v^*}^*) > \epsilon_1) \leq C_2 e^{-n^2 \epsilon_2}.$$

Since $|\mathcal{T}(X) - (v^*)^3| \leq 1$,

$$(I) \leq \mathbb{Q}_n(\delta_\square(X, \mathcal{F}_{v^*}^*) > \epsilon_1) \leq C_2 e^{-n^2 \epsilon_2}.$$

To estimate (II), for any $X \in \{\delta_\square(X, \mathcal{F}_{v^*}^*) \leq \epsilon_1\}$, let the function $f_X^* \in \mathcal{F}_{v^*}^*$ be such that $\delta_\square(X, f_X^*) \leq \epsilon_1$. Note that $\mathcal{T}(f_X^*) = (v^*)^3$ by definition. By Lipschitz continuity of the mapping $f \mapsto \mathcal{T}(f)$ under the cut distance metric δ_\square [3, Theorem 3.7],

$$|\mathcal{T}(X) - (v^*)^3| = |\mathcal{T}(X) - \mathcal{T}(f_X^*)| \leq C_1 \delta_\square(X, f_X^*) \leq C_1 \epsilon_1.$$

So

$$\begin{aligned}
(II) &= \int_{\{\delta_\square(X, \mathcal{F}_{v^*}^*) \leq \epsilon_1\}} |\mathcal{T}(X) - (v^*)^3| d\mathbb{Q}_n(X) \\
&\leq C_1 \epsilon_1 \mathbb{Q}_n(\delta_\square(X, \mathcal{F}_{v^*}^*) \leq \epsilon_1) \\
&\leq C_1 \epsilon_1.
\end{aligned}$$

Hence,

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{Q}_n} |\mathcal{T}(X) - (v^*)^3| \leq \lim_{n \rightarrow \infty} C_2 e^{-n^2 \epsilon_2} + C_1 \epsilon_1 = C_1 \epsilon_1.$$

Since ϵ_1 is arbitrary, (2.27) follows.

If (q, v^*) belongs to the replica symmetric phase, we have by Theorem 2.5 that $\mathcal{F}_{v^*}^*$ consists uniquely of the constant function $f^*(x, y) \equiv v^*$. Then since $\mathcal{E}(f^*) = v^*$, the above proof follows identically to yield that

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{Q}_n} |\mathcal{E}(X) - v^*| \leq \lim_{n \rightarrow \infty} C_2 e^{-n^2 \epsilon_2} + C \epsilon_1 = C \epsilon_1.$$

■

Lemma A.2. *Let \mathcal{S}_α be defined in Definition 2.10. Then $\mathcal{S}_\alpha \supset \mathcal{S}_{\alpha'}$ for $0 < \alpha < \alpha'$.*

Proof. Denote $\phi_p^\alpha(x) = \phi_p(x^{1/3\alpha})$ and let $\hat{\phi}_p^\alpha(x)$ be the convex minorant of $\phi_p^\alpha(x)$. Then

$$\phi_p^{\alpha'}(x) = \phi_p(x^{1/3\alpha'}) = \phi_p((x^{\alpha/\alpha'})^{1/3\alpha}) = \phi_p^\alpha(x^{\alpha/\alpha'}).$$

Define $\eta(x) = \hat{\phi}_p^{\alpha'}(x^{\alpha'/\alpha})$. Let K be the set where $\eta(x) = \phi_p^{\alpha'}(x^{\alpha'/\alpha})$ for $x \in K$. Then

$$\eta(x) \leq \phi_p^{\alpha'}(x^{\alpha'/\alpha}) = \phi_p^\alpha(x)$$

with equality occurring iff $x \in K$. (The interpretation of K is that $t^{3\alpha} \in K$ if and only if (p, t) satisfies the minorant condition with α' .) Since $\frac{\alpha'}{\alpha} > 1$, the function $\eta(x)$ is convex and is less than $\phi_p^\alpha(x)$, hence it must be less than the convex minorant, $\eta(x) \leq \hat{\phi}_p^\alpha(x)$. For $x \in K$,

$$\phi_p^\alpha(x) = \eta(x) \leq \hat{\phi}_p^\alpha(x) \leq \phi_p^\alpha(x)$$

so $(x, \phi_p^\alpha(x))$ lies on the convex minorant $\hat{\phi}_p^\alpha(x)$ for all $x \in K$. Hence, if (p, t) satisfying the minorant condition with α' , then $t^{3\alpha} \in K$ and $(t^{3\alpha}, \phi_p(t))$ lies on the convex minorant $\hat{\phi}_p^\alpha(x)$, implying that (p, t) satisfies the minorant condition with α .

Now let (p, t) satisfy the minorant condition with α' , and suppose that $\frac{\beta'}{6}$ is a subdifferential of $\hat{\phi}_p^{\alpha'}(x)$ at the point $t^{3\alpha'}$ such that $\sup[\frac{\beta'}{6}u^{3\alpha} - \phi_p(u)]$ is uniquely maximized at t . According to the arguments in the proof of Lemma 2.9, this means that the subgradient line

$$\ell_{\alpha'}(x) := \frac{\beta'}{6}(x - t^{3\alpha'}) - \phi_p(t)$$

lies below $\phi_p^{\alpha'}(x)$ and touches it at exactly one point $t^{3\alpha'}$. Let $\nu(x) = \ell_{\alpha'}(x^{\alpha'/\alpha})$. We have that $\nu(t^{3\alpha}) = \phi_p^{\alpha'}(t^{3\alpha'}) = \phi_p^\alpha(t^{3\alpha})$ and $\nu'(t^{3\alpha}) = \frac{\beta'}{6} \frac{\alpha'}{\alpha} t^{3(\alpha' - \alpha)}$. Since $\frac{\alpha'}{\alpha} > 1$, $\nu(x)$ is convex, and the line

$$\ell_\alpha(x) := \frac{\beta}{6}(x - t^{3\alpha}) - \phi_p(t),$$

where $\frac{\beta}{6} = \nu'(t^{3\alpha})$, is tangent to $\nu(x)$ at the point $t^{3\alpha}$ and lies below $\nu(x)$. For $x \neq t^{3\alpha}$,

$$\nu(x) = \ell_{\alpha'}(x^{\alpha'/\alpha}) < \phi_p^{\alpha'}(x^{\alpha'/\alpha}) = \phi_p^\alpha(x),$$

so $\ell_\alpha(x)$ lies below $\phi_p^\alpha(x)$ and touches it at exactly one point $t^{3\alpha}$. Moreover, since $\nu(x)$ is a convex function less than $\phi_p^\alpha(x)$, we have $\hat{\phi}_p^\alpha(x) \geq \nu(x) \geq \ell_\alpha(x)$. So, $\frac{\beta}{6}$ is a subdifferential of $\hat{\phi}_p^\alpha(x)$ and $\sup[\frac{\beta}{6}u^{3\alpha} - \phi_p(u)]$ is uniquely maximized at t .

The proof is complete. ■

Remark A.3. We note an interesting connection between the subdifferentials β' and β in the above proof. If $\phi_p(t)$ is differentiable at t , then (2.30) explicitly specifies the relationship between the subdifferentials:

$$\beta = \frac{2\phi_p'(t)}{\alpha t^{3\alpha-1}} = \frac{2\phi_p'(t)}{\alpha' t^{3\alpha'-1}} \frac{\alpha' t^{3\alpha'-1}}{\alpha t^{3\alpha-1}} = \frac{\alpha'}{\alpha} t^{3(\alpha' - \alpha)} \beta'.$$

This is consistent with the derivation in the above proof.

Proof of Proposition 4.2.

Proof. The proof is identical to that of Theorem 2.5 with a few obvious modifications. For $u \in J$, suppose $f \in A_0 \cap \partial\mathcal{W}_u$. Similarly to (2.26),

$$\mathcal{H}(f) - \mathcal{I}(f) = \frac{\beta}{6}u^{3\alpha} - \mathcal{I}_q(f) - \frac{1}{2}\log(1-q)$$

so

$$\begin{aligned} \sup_{f \in A_J} [\mathcal{H}(f) - \mathcal{I}(f)] &= \sup_{u \in J} \sup_{f \in A_0 \cap \partial\mathcal{W}_u} [\mathcal{H}(f) - \mathcal{I}(f)] \\ &\leq \sup_{u \in J} \left[\frac{\beta}{6}u^{3\alpha} - \inf_{f \in A_0 \cap \partial\mathcal{W}_u} [\mathcal{I}_q(f)] - \frac{1}{2}\log(1-q) \right] \\ &= \frac{\beta}{6}(v^*)^{3\alpha} - \mathcal{I}_q(f^*) - \frac{1}{2}\log(1-q) \end{aligned}$$

where now

$$v^* := \operatorname{argsup}_{u \in J} \left[\frac{\beta}{6}(v^*)^{3\alpha} - \inf_{f \in A_0 \cap \partial\mathcal{W}_u} [\mathcal{I}_q(f)] - \frac{1}{2}\log(1-q) \right]$$

and f^* is any function in $\mathcal{F}_{v^*}^*$. The supremum $\sup_{f \in A_J} [\mathcal{H}(f) - \mathcal{I}(f)]$ is attained on the set $\mathcal{F}_{v^*}^*$. This concludes the proof of (4.5). The proof of the second part of Proposition 4.2 follows identically to the proof in Theorem 2.5 and is omitted. \blacksquare

We collect here some basic asymptotic properties of some probabilities of interest.

Lemma A.4. *Let A_J be defined in (4.4).*

(i) *Suppose $p \in J$. Let $\kappa_{n,p,A_J} := \frac{1}{n^2} \log \sum_{X \in A_J} e^{h_p \mathcal{E}(X)}$ and $\kappa_{n,p} = \frac{1}{n^2} \log \sum_{X \in \Omega_n} e^{h_p \mathcal{E}(X)}$ be the normalizing constants for \mathbb{P}_{n,p,A_J} and $\mathbb{P}_{n,p}$, respectively. Then*

$$\lim_{n \rightarrow \infty} \kappa_{n,p,A_J} = \lim_{n \rightarrow \infty} \kappa_{n,p} = -\frac{1}{2} \log(1-p).$$

Moreover, $\mathbb{P}_{n,p}(A_J) \rightarrow 1$ as $n \rightarrow \infty$.

(ii) *Assume that $t \in J^\circ$ is in the interior of J . Also assume that $\delta_\square(\tilde{F}^*, A_0^c) > 0$, where $\tilde{F}^* \subset \mathcal{W}$ is the set of minimizers of the LDP rate function $\inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]$. Then*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}(\mathcal{W}_t \cap A_J^c) < - \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)].$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p,A_J}(\mathcal{W}_t) = - \inf_{f \in \mathcal{W}_t \cap A_J} [\mathcal{I}_p(f)] = - \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)].$$

Proof. (i) From Theorem 4.2, since $p = \operatorname{argsup}_{0 \leq u \leq 1} [V(u)]$ and $p \in J$,

$$\lim_{n \rightarrow \infty} \kappa_{n,p,A_J^c} = \sup_{u \in J} [V(u)] = \sup_{0 \leq u \leq 1} [V(u)] = \lim_{n \rightarrow \infty} \kappa_{n,p}$$

From (2.1), $\lim_{n \rightarrow \infty} \kappa_{n,p} = \frac{1}{2} \log(1-p)$. It follows directly that $\mathbb{P}_{n,p}(A_J) \rightarrow 1$ as $n \rightarrow \infty$.

(ii) Since $\delta_\square(\tilde{F}^*, A_0^c) > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}(\mathcal{W}_t \cap A_0^c) = - \inf_{f \in \mathcal{W}_t \cap A_0^c} [\mathcal{I}_p(f)] < - \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)].$$

If r is the smallest value larger than t in J^C , then $-\inf_{f \in \mathcal{W}_t \cap A_J^c} [I_p(f)] < -\inf_{f \in \mathcal{W}_t} [I_p(f)]$, so

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}(\mathcal{W}_t \cap A_J^c) \leq -\max \left\{ \inf_{f \in \mathcal{W}_t \cap A_J^c} [I_p(f)], \inf_{f \in \mathcal{W}_t \cap A_J} [I_p(f)] \right\} < \inf_{f \in \mathcal{W}_t} [I_p(f)].$$

Also,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p,A_J}(\mathcal{W}_t) &= \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}(\mathcal{W}_t \cap A_J) - \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}(A_J) \\ &= -\inf_{f \in \mathcal{W}_t \cap A_J} [\mathcal{I}_p(f)] = -\inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]. \end{aligned}$$

The last equality follows since $t \in J$ and $\inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)] = \inf_{f \in \partial \mathcal{W}_t} [\mathcal{I}_p(f)]$ (Theorem 4.3 in [8]). \blacksquare

Proof of Corollary 4.3.

Proof. Denote

$$\hat{q}_{n,A_J} := \mathbf{1}_{\mathcal{W}_t} \frac{d\tilde{\mathbb{P}}_n}{d\tilde{\mathbb{Q}}_n}.$$

(For brevity, we drop the sub/superscripts p, h, β, α, A_J if no ambiguity arises; thus, \mathbb{Q}_n denotes the Gibbs measure $\mathbb{Q}_n^{h,\beta,\alpha}$ and $\tilde{\mathbb{Q}}_n$ denotes the conditional measure $\tilde{\mathbb{Q}}_n^{h,\beta,\alpha}$, and similarly for $\mathbb{P}_n, \tilde{\mathbb{P}}_n$.) Under $\tilde{\mathbb{Q}}_n$, the second moment is

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbb{Q}}_n}[\hat{q}_{n,A_J}^2] &= \frac{\mathbb{E}_{\mathbb{P}_n}[\mathbf{1}_{\mathcal{W}_t \cap A_J} \frac{d\tilde{\mathbb{P}}_n}{d\tilde{\mathbb{Q}}_n}]}{\mathbb{P}_n(A_J)} \\ &= (\mathbb{P}_n(A_J))^{-1} \mathbb{E}_{\mathbb{P}_n}[\mathbf{1}_{\mathcal{W}_t \cap A_J} e^{n^2(-\mathcal{H}(X) + \frac{h_p}{2}\mathcal{E}(X))}] e^{n^2(\tilde{\psi}_{n,A_J} - \kappa_{n,p,A_J})} \end{aligned}$$

where $\kappa_{n,p,A_J} = \frac{1}{n^2} \log \sum_{X \in A_J} e^{\frac{h_p}{2}\mathcal{E}(X)}$ is the normalizing constant for \mathbb{P}_{n,p,A_J} .

With $h = h_p$, we apply the Laplace principle and Lemma A.4(i),

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{E}_{\tilde{\mathbb{Q}}_n}[\hat{q}_{n,A_J}^2] \\ &= -\inf_{f \in \mathcal{W}_t \cap A_J} \left[\mathcal{I}_p(f) + \frac{\beta}{6} \mathcal{T}(f)^\alpha + \frac{h - h_p}{2} \mathcal{E}(f) \right] + \lim_{n \rightarrow \infty} \tilde{\psi}_{n,A_J} + \frac{1}{2} \log(1 - p) \\ &= -\inf_{f \in \mathcal{W}_t \cap A_J} \left[\mathcal{I}_p(f) + \frac{\beta}{6} \mathcal{T}(f)^\alpha + \frac{h - h_p}{2} \mathcal{E}(f) \right] + \frac{\beta}{6} t^{3\alpha} - \inf_{f \in \partial \mathcal{W}_t} [\mathcal{I}_p(f)] \\ &\geq -\inf_{f \in \mathcal{W}_t \cap A_J} [\mathcal{I}_p(f)] - \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)] = -2 \inf_{f \in \mathcal{W}_t} [\mathcal{I}_p(f)]. \end{aligned}$$

Hence the importance sampling scheme is asymptotically optimal. \blacksquare

REFERENCES

- [1] S. Bhamidi, G. Bresler, and A. Sly, *Mixing time of exponential random graphs*, Ann. Appl. Probab. **21** (2011), no. 6, 2146–2170. MR2895412
- [2] J. Blanchet and P. Glynn, *Efficient rare-event simulation for the maximum of heavy-tailed random walks*, The Annals of Applied Probability **18** (2008), no. 4, 1351–1378.
- [3] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztegombi, *Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing*, Adv. Math. **219** (2008), no. 6, 1801–1851. MR2455626 (2009m:05161)
- [4] J. A. Bucklew, *Introduction to rare event simulation*, Springer Series in Statistics, Springer-Verlag, New York, 2004. MR2045385 (2005e:62001)

- [5] S. Chatterjee and P. Diaconis, *Estimating and understanding exponential random graph models*, arXiv preprint arXiv:1102.2650 (2011).
- [6] S. Chatterjee, *The missing log in large deviations for triangle counts*, Random Structures Algorithms **40** (2012), no. 4, 437–451. MR2925306
- [7] S. Chatterjee and P. S. Dey, *Applications of Stein’s method for concentration inequalities*, Ann. Probab. **38** (2010), no. 6, 2443–2485. MR2683635 (2012f:60073)
- [8] S. Chatterjee and S. R. S. Varadhan, *The large deviation principle for the Erdős-Rényi random graph*, European J. Combin. **32** (2011), no. 7, 1000–1017. MR2825532 (2012m:60067)
- [9] B. DeMarco and J. Kahn, *Upper tails for triangles*, Random Structures Algorithms **40** (2012), no. 4, 452–459. MR2925307
- [10] P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stochastics: An International Journal of Probability and Stochastic Processes **76** (2004), no. 6, 481–508.
- [11] P. Glasserman and Y. Wang, *Counterexamples in importance sampling for large deviations probabilities*, Ann. Appl. Probab. **7** (1997), no. 3, 731–746. MR1459268 (98b:60053)
- [12] S. Juneja and P. Shahabuddin, *Rare event simulation techniques: An introduction and recent advances*, Simulation, Handbooks in Operations Research and Management Science (2006), 291–350.
- [13] J. H. Kim and V. H. Vu, *Divide and conquer martingales and the number of triangles in a random graph*, Random Structures Algorithms **24** (2004), no. 2, 166–174. MR2035874 (2005d:05135)
- [14] L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957. MR2274085 (2007m:05132)
- [15] L. Lovász and B. Szegedy, *Szemerédi’s lemma for the analyst*, Geom. Funct. Anal. **17** (2007), no. 1, 252–270. MR2306658 (2008a:05129)
- [16] E. Lubetzky and Y. Zhao, *On replica symmetry of large deviations in random graphs*, arXiv preprint arXiv:1210.7013 (2012).
- [17] C. Radin and M. Yin, *Phase transitions in exponential random graphs*, arXiv preprint arXiv:1108.0649 (2011).
- [18] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, *An introduction to exponential random graph (p^*) models for social networks*, Social networks **29** (2007), no. 2, 173–191.
- [19] G. Robins, T. Snijders, P. Wang, M. Handcock, and P. Pattison, *Recent developments in exponential random graph (p^*) models for social networks*, Social networks **29** (2007), no. 2, 192–215.
- [20] G. Rubino and B. Tuffin, *Rare event simulation using monte carlo methods*, Wiley Online Library, 2009.
- [21] M. Yin, *A cluster expansion approach to exponential random graph models*, Journal of Statistical Mechanics: Theory and Experiment **2012** (2012), no. 05, P05004.
- [22] M. Yin, *Critical phenomena in exponential random graphs*, arXiv preprint arXiv:1208.2992 (2012).

¹DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, 304 HANES HALL, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599

²DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, 330 HANES HALL, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599

³STATISTICAL AND APPLIED MATHEMATICAL SCIENCES INSTITUTE, 19 T.W. ALEXANDER DRIVE, P.O. BOX 14006, RESEARCH TRIANGLE PARK, NC 27709, USA.

MATHEMATICS DEPARTMENT, DUKE UNIVERSITY, BOX 90320, DURHAM, NORTH CAROLINA, 27708, USA

E-mail address: bhamidi@email.unc.edu, hannig@email.unc.edu, leecy@email.unc.edu, nolen@math.duke.edu